


```

:      ORGANISM: Nicotiana tabacum/Cladosporium fulvum
:
:      STRAIN: Cladosporium fulvum race 2,5
:
:      IMMEDIATE SOURCE:
:
:      CLONE: SP:AVR4
:
US-08-945-983-3

```

Query Match	31.28;	Score 61.4;	DB 4;	Length 484;
Best Local Similarity	67.78;	Pred. No. 2e-10;		
Matches 86; Conservative	0;	Mismatches 41;	Indels 0;	Gaps 0;

QY	130	actgagaa	136
Db	123	CATGCAA	129

RESULT 2
US-08-945-983-5/c

```

: GENERAL INFORMATION:
: APPLICANT: Thomas, Colwyn M
: APPLICANT: Balint-Kurtli, Peter J
: APPLICANT: Jones, David A
: APPLICANT: Jones, Jonathan DG
: TITLE OF INVENTION: Plant pathogen resistance genes and uses
: TITLE OF INVENTION: thereof
: NUMBER OF SEQUENCES: 9
: CORRESPONDENCE ADDRESS:

```

```

1  COMPUTER READABLE FORM:
2  MEDIUM TYPE: Floppy disk
3  * OPERATOR: IBM PC compatible
4  * OPERATING SYSTEM: PC-DOS/MS-DOS
5  SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
6  CURRENT APPLICATION DATA:
7  APPLICATION NUMBER: US/08/945,983
8  FILING DATE: 12-NOV-1997

```

1 PRIOR APPLICATION DATA:
 2 APPLICATION NUMBER: PCT/GB96/01155
 3 FILING DATE: 13-MAY-1996
 4 PRIOR APPLICATION DATA:
 5 APPLICATION NUMBER: GB 9509575.8
 6 FILING DATE: 11-MAY-1995
 7 ATTORNEY/AGENT INFORMATION:
 8

```

; INFORMATION FOR SEQ ID NO:
;
; SEQUENCE CHARACTERISTICS:
;
;     LENGTH: 484 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: double
;     TOPOLOGY: linear
;
US-08-945-983-5

```

Query Match	31.2%;	Score 61.4;	DB 4;	Length 48;
Best Local Similarity	67.7%;	Pred. No. 2e-10;		
Matches 86; Conservative	0;	Mismatches 41;	Indels 0;	Gaps 0;

Qy	10	actgtggtttcttgctcttcttcacagcttccacactcttcctctgctctctctcttc	69
Db	482	CGATGGGATTTTCTTCCTTTACAAATTTGCCTTCATTTCTCTGCTGCTACACTTCCT	423
Qy	70	ttctctcttgactctcactcctctgctgctgctgctccagaagccgctgtgccaatcact	128
Db	422	TATTTCTATATATATCCACACTCTTGGCGTGCACAAAGCCCAAAATCTACACATATCAACC	365
Qy	130	actgtcaa	136
Db	362	CATGCCAA	356

RESULT 3
US-08-181-271A-9
; Sequence 9, Application US/08181271A
; Patent No. 5614395

1 GENERAL INFORMATION:
2
3 APPLICANT: Ryals, John A.
4 APPLICANT: Alexander, Danny C.
5 APPLICANT: Beck, James J.
6 APPLICANT: Duesing, John H.
7 APPLICANT: Friedrich, Leslie B.
8 APPLICANT: Goodman, Robert M.
9 APPLICANT: Harms, Christian
10 APPLICANT: Melns, Jr., Frederick
11 APPLICANT: Montoya, Alice
12 APPLICANT: Moyer, Mary B.
13 APPLICANT: Neuhaus, Jean-Marc
14 APPLICANT: Payne, George B.
15 APPLICANT: Sperison, Christoph
16 APPLICANT: Stinson, Jeffrey R.
17 APPLICANT: Uknes, Scott J.

```

:
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
:

```

```

1      PRIOR APPLICATION DATA:
2      APPLICATION NUMBER:  US 08/093,301
3      FILING DATE:  16-JUL-1993
4      PRIOR APPLICATION DATA:
5      APPLICATION NUMBER:  US 07/937,197
6      FILING DATE:  6-NOV-1992
7      PRIOR APPLICATION DATA:
8      APPLICATION NUMBER:  US 07/678,378
9      FILING DATE:  1-APR-1991
10     PRIOR APPLICATION DATA:
11     APPLICATION NUMBER:  US 07/305,566
12     FILING DATE:  6-FEB-1989
13     PRIOR APPLICATION DATA:
14     APPLICATION NUMBER:  US 07/165,667

```


FILED DATE: 20-JUN-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/329,018
FILING DATE: 24-MAR-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/045,957
FILING DATE: 12-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: S-19825/P1/CGC 1727
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919)541-8614
TELEFAX: (919)541-8689
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 809 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-455-416-9

Query Match 29.3%; Score 57.8; DB 1; Length 809;
Best Local Similarity 76.3%; Pred. No. 3.3e-09;
Matches 71; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY 9 gccatggttcgtcttcctcagcttcacatcttcctctgtctactctct 68
DB 27 gtcacggattgtctctttcacacattgccttcattcttctgtctacacttctc 86
QY 69 ctttccttgatctcctcactcgtcgctgc 101
DB 87 ttattcttgatattccacactcttcgctgc 119

RESULT 9
US-08-455-244-9
Sequence 9, Application US/08455244
Patent No. 5789214
GENERAL INFORMATION:
APPLICANT: Ryals, John A.
APPLICANT: Alexander, Danny C.
APPLICANT: Beck, James J.
APPLICANT: Duesing, John H.
APPLICANT: Friedrich, Leslie B.
APPLICANT: Goodman, Robert M.
APPLICANT: Harms, Christian
APPLICANT: Meins, Jr., Frederick
APPLICANT: Montoya, Alice
APPLICANT: Moyer, Mary B.
APPLICANT: Neuhaus, Jean-Marc
APPLICANT: Payne, George B.
APPLICANT: Spelison, Christoph
APPLICANT: Stinson, Jeffrey R.
APPLICANT: Uknes, Scott J.
APPLICANT: Ward, Eric R.
APPLICANT: Williams, Shericea C.
TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF
NUMBER OF SEQUENCES: 106
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: New York
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/455,244
FILING DATE: 31-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/181,271
FILING DATE: 13-JAN-94
APPLICATION NUMBER: US 08/093,301
FILING DATE: 16-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/937,197
FILING DATE: 6-NOV-1992
APPLICATION NUMBER: US 07/678,378
FILING DATE: 1-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/305,566
FILING DATE: 6-FEB-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/165,667
FILING DATE: 8-MAR-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/042,847
FILING DATE: 6-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/632,441
FILING DATE: 21-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/425,504
FILING DATE: 20-OCT-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/848,506
FILING DATE: 6-MAR-1992
APPLICATION NUMBER: US 07/768,122
FILING DATE: 27-SEP-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/580,431
FILING DATE: 7-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/368,672
FILING DATE: 20-JUN-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/329,018
FILING DATE: 24-MAR-1989
APPLICATION NUMBER: US 08/045,957
FILING DATE: 12-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: S-19825/P1/CGC 1727
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919)541-8614
TELEFAX: (919)541-8689
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 809 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-455-244-9

Query Match 29.3%; Score 57.8; DB 1; Length 809;
Best Local Similarity 76.3%; Pred. No. 3.3e-09;
Matches 71; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY 9 gccatggttcgtcttcctcagcttcacatcttcctctgtctactctct 68
DB 27 gtcacggattgtctctttcacacattgccttcattcttctgtctacacttctc 86

CORRESPONDENCE ADDRESS:
 ADDRESSEE: CIBA-GEIGY Corporation
 STREET: 7 Skyline Drive
 CITY: Hawthorne
 STATE: New York
 COUNTRY: USA
 ZIP: 10532
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/457,364
 FILING DATE: 31-MAY-1995
 CLASSIFICATION: 800
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/181,271
 FILING DATE: 13-JUN-94
 APPLICATION NUMBER: US 08/093,301
 FILING DATE: 16-JUL-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/937,197
 FILING DATE: 6-NOV-1992
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/678,378
 FILING DATE: 1-APR-1991
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/305,566
 FILING DATE: 6-FEB-1989
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/165,667
 FILING DATE: 8-MAR-1988
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/042,847
 FILING DATE: 6-APR-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/632,441
 FILING DATE: 21-DEC-1990
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/425,504
 FILING DATE: 20-OCT-1989
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/848,506
 FILING DATE: 6-MAR-1992
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/768,122
 FILING DATE: 27-SEP-1991
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/580,431
 FILING DATE: 7-SEP-1990
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/368,672
 FILING DATE: 20-JUN-1989
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/329,018
 FILING DATE: 24-MAR-1989
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/045,957
 FILING DATE: 12-APR-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Elmer, James Scott
 REGISTRATION NUMBER: 36,129
 REFERENCE/DOCKET NUMBER: S-19825/P1/CGC 1727
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (919)541-8614
 TELEFAX: (919)541-8689
 INFORMATION FOR SEQ ID NO: 9:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 809 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear

MOLECULE TYPE: CDNA
 US-08-457-364-9
 Query Match 29.3%; Score 57.8; DB 2; Length 809;
 Best Local Similarity 76.3%; Pred. No. 3.3e-09;
 Matches 71; Conservative 0; Mismatches 22; Indels 0; Gaps 0;
 QY 9 gccatgggttctgtcttcttcacgttcacattcttccttctgtgttactcttct 68
 Db 27 gcatgggattgtcttcttcttgcacattgccttcttcttcttcttcttct 86
 QY 69 cttctctgtatctctcactcttgcgtgac 101
 Db 87 ttattcttgatattatccactcttgccgtgcc 119
 RESULT 12
 US-08-456-262-9
 Sequence 9, Application US/08456262
 Patent No. 5851766
 GENERAL INFORMATION:
 APPLICANT: Ryals, John A.
 APPLICANT: Alexander, Danny C.
 APPLICANT: Beck, James J.
 APPLICANT: Duesing, John H.
 APPLICANT: Friedrich, Leslie B.
 APPLICANT: Goodman, Robert M.
 APPLICANT: Harms, Christian
 APPLICANT: Meins, Jr., Frederick
 APPLICANT: Montoya, Alice
 APPLICANT: Moyer, Mary B.
 APPLICANT: Neuhaus, Jean-Marc
 APPLICANT: Payne, George B.
 APPLICANT: Sperison, Christoph
 APPLICANT: Stinson, Jeffrey R.
 APPLICANT: Unnes, Scott J.
 APPLICANT: Ward, Eric R.
 TITLE OF INVENTION: Williams, Shericea C.
 TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
 NUMBER OF SEQUENCES: 106
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: CIBA-GEIGY Corporation
 STREET: 7 Skyline Drive
 CITY: Hawthorne
 STATE: New York
 COUNTRY: USA
 ZIP: 10532
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/456,262
 FILING DATE: 31-MAY-1995
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/181,271
 FILING DATE: 13-JAN-94
 APPLICATION NUMBER: US 08/093,301
 FILING DATE: 16-JUL-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/937,197
 FILING DATE: 6-NOV-1992
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/678,378
 FILING DATE: 1-APR-1991
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/305,566
 FILING DATE: 6-FEB-1989
 PRIOR APPLICATION DATA:

Patent No. 5942662
GENERAL INFORMATION:
APPLICANT: Ryals, John A.
APPLICANT: Harms, Christian
APPLICANT: Friedrich, Leslie
APPLICANT: Beck, James
APPLICANT: Uknes, Scott
APPLICANT: Ward, Eric
TITLE OF INVENTION: INDUCIBLE HERBICIDE RESISTANCE
NUMBER OF SEQUENCES: 111
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 594262artis Corporation
STREET: 3054 Cornwells Road, P.O. Box 12257
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/971,217
FILING DATE:
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/457,364
FILING DATE: 31-MAY-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/181,271
FILING DATE: 13-JAN-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/093,301
FILING DATE: 16-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/937,197
FILING DATE: 6-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/678,378
FILING DATE: 1-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/305,566
FILING DATE: 6-FEB-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/165,667
FILING DATE: 8-MAR-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/042,847
FILING DATE: 6-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/632,441
FILING DATE: 21-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/425,504
FILING DATE: 20-OCT-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/848,506
FILING DATE: 6-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/768,122
FILING DATE: 27-SEP-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/580,431
FILING DATE: 7-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/368,672
FILING DATE: 20-JUN-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/329,018
FILING DATE: 24-MAR-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/045,957

FILING DATE: 12-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38, 241
REFERENCE/DOCKET NUMBER: S-19825/P1/CGC 1727/DIV5/CONF
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919)541-8587
TELEFAX: (919)541-8689
INFORMATION FOR SEQ. ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 809 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-971-217-9
Query Match 29.3%; Score 57.8; DB 2; Length 809;
Best Local Similarity 76.3%; Pred. No. 3.3e-09;
Matches 71; Conservative 0; Mismatches 22; Indels 0; Gaps 0;
QY 9 gccatgggttcgctcttcctcctcagcttcacatcttctctcttgatgctacattctt 68
Db 27 gtcattggatttctctcttccacattgcttccattcttctctctctctctctctc 86
QY 69 ctcttccttgatctctcctcactcttgccgtgccc 101
Db 87 ttattcttattatattatccacattcttgcctgccc 119

Search completed: May 12, 2002, 20:14:37
Job time: 7194 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 12, 2002, 20:20:39 ; Search time 333.08 Seconds
(without alignments)

1015.468 Million cell updates/sec

Title: US-09-554-024-5

Perfect score: 197

Sequence: 1 ggcgtgcagccatgggttc.....gcagcgtgcagcagatcc 197

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_032802.*

1: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1980.DAT.*

2: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1981.DAT.*

3: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1982.DAT.*

4: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1983.DAT.*

5: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1984.DAT.*

6: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1985.DAT.*

7: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1986.DAT.*

8: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1987.DAT.*

9: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1988.DAT.*

10: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1989.DAT.*

11: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1990.DAT.*

12: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1991.DAT.*

13: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1992.DAT.*

14: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1993.DAT.*

15: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1994.DAT.*

16: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1995.DAT.*

17: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1996.DAT.*

18: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1997.DAT.*

19: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT.*

20: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT.*

21: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT.*

22: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT.*

23: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT.*

24: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	197	100.0	197	20	AAK59774
2	117.4	59.6	211	20	AAK29729
3	103	52.3	106	20	AAK59773
4	103	52.3	106	20	AAK29728
5	101.6	51.6	120	21	AAZ45694
6	96	48.7	98	20	AAK59772
7	93.4	47.4	261	21	AAZ45693
8	73.4	37.3	75	20	AAK29730
9	67.4	34.2	72	20	AAK29731

10	63	32.0	63	20	AAK59771	CDNA encoding a th
11	61.4	31.2	484	17	AAK06310	Chimeric avirulence
12	61.4	31.2	484	17	AAK47876	C. fulvum avirulen
13	58.4	29.6	107	22	AAK10075	PR1 leader sequen
14	57.8	29.3	809	11	AAK06180	PR-1a CDNA cloned
15	57.8	29.3	809	20	AAK62805	Tobacco PR-1a gene
16	57.8	29.3	809	20	AAK72997	PR-1a protein enco
17	57.8	29.3	1260	11	AAK04693	Promoter of the in
18	57.8	29.3	1363	10	AAK91026	Genomic sequence o
19	57.8	29.3	2038	10	AAK90367	Fragment of tobacc
20	57.8	29.3	2038	11	AAK03663	Fragment of tobacc
21	57.8	29.3	2038	20	AAK62787	Tobacco PR-1a gene
22	57.8	29.3	2038	20	AAK72999	Tobacco PR-1a gene
23	57.8	29.3	216	15	AAK63461	SPR1-Shiva I fusio
24	56.8	28.8	177	16	AAK09148	Chimeric Avr9 gene
25	56.8	28.8	177	17	AAK06305	Chimeric Avr9 gene
26	56.4	28.6	90	20	AAK06760	C-terminus of PR-1
27	56.4	28.6	90	22	AAK12812	Nicotiana tabacum
28	56.4	28.6	90	22	AAK55702	Tobacco PR-1a sign
29	56.4	28.6	213	15	AAK63462	SPR1-SR37 fusion p
30	52	26.4	772	10	AAK91019	Nicotiana tabacum
31	51.6	26.2	110	22	AAK12810	Nicotiana tabacum
32	49.8	25.3	751	20	AAK80793	Synthetic ORF5 gen
33	49.8	25.3	771	11	AAK06181	PR-1b CDNA cloned
34	49.8	25.3	771	20	AAK62806	Tobacco PR-1b gene
35	49.8	25.3	771	20	AAK72998	PR-1b protein enco
36	48.4	24.6	91	20	AAK80795	Primer-2 for const
37	48.4	24.6	102	22	AAK12811	Nicotiana tabacum
38	48.4	24.6	207	21	AAK28519	PCR11 DNA coding s
39	48.4	24.6	242	11	AAK06552	sal10 fused to PRL
40	47.2	24.0	656	11	AAK06182	Tobacco PR-1c gene
41	47.2	24.0	656	20	AAK62807	Tobacco PR-1c gene
42	47.2	24.0	656	20	AAK72999	PR-1c protein enco
43	43.2	21.9	735	10	AAK91020	Nicotiana tabacum
44	42.6	21.6	169	19	AAK50383	Linked TMV leader
45	42.4	21.5	123219	23	AAK88703	Human DNA sequence

ALIGNMENTS

AAK59774	1	AAK59774 standard; CDNA; 197 BP.
AC	AAK59774:	
DT	26-JUL-1999 (first entry)	
XX	CDNA encoding a thanatine fusion product of the invention.	
DE		
XX	Thranatine; fungal disease; bacterial disease; Cercospora beticola;	
KW	Cladosporium herbarum; Fusarium cuneiform; F. graminearum;	
KW	Phytophthora cinnamomi; selection marker; plant transformation;	
KW	herbicide resistance; PR-1a gene; tobacco; ss.	
OS	Synthetic.	
XX		
PN	FR2770853-AL.	
PD	14-MAY-1999.	
XX		
PF	07-NOV-1997; 97FR-0014263.	
XX		
PR	07-NOV-1997; 97FR-0014263.	
XX		
PA	(RHON) RHONE-FOULENC AGROCHIMIE.	
PI	Derose R. Freyssinet G, Hoffmann J;	
XX		
DR	WPI: 1999-315645/27.	
XX	P-PSDB; AAY15466.	

PT	New nucleic acid encoding thanatine useful as a selection marker for transformation of plants
PS	Claim 8; Page 14-15; 24pp; French.
XX	The specification describes a nucleic acid sequence containing the sequence that encodes thanatine. Plants transformed with this nucleic acid sequence are resistant to fungal and bacterial diseases, specifically those caused by <i>Cercospora beticola</i> , <i>Cladosporium herbarum</i> , <i>Fusarium eumorium</i> , <i>F. graminearum</i> and <i>Phytophthora cinnamomi</i> . The nucleic acid sequence may also be used as a selection marker for transformation of plants with other coding sequences, e.g. those that impart resistance to herbicides. The present sequence encodes a thanatine fusion product of the invention, comprising the signal peptide of the PR-1a gene of tobacco.
CC	Sequence 197 BP; 32 A; 56 C; 49 G; 60 T; 0 other;
YY	Query Match 100.0%; Score 197; DB 20; Best Local Similarity 100.0%; Pred. NO. 6.4e-51; Matches 197; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Dd	1 ggcgtcagacgcaatgggttcgcttctctccaaattccactttctcttggtcta 60 1 ggcgtcagacgcaatgggttcgcttctctccaaattccactttctcttggtcta 60
OY	61 ctctctctcttctctcttgatctctgaacctctctcgccggttcacaagaaccagtgc 120 61 ctctctctcttctctcttgatctctgaacctctctcgccggttcacaagaaccagtgc 120
Dd	121 caatcatctactgcacaacaggagactggttaagtccagaagatgtgagctcggcgagcg 180 121 caatcatctactgcacaacaggagactggttaagtccagaagatgtgagctcggcgagcg 180
OY	181 aacgtctcgacggaatcc 197 181 aacgtctcgacggaatcc 197
Dd	RESULT 2 AAAX29729 standard; DNA; 211 BP. AAAX29729: 22-JUN-1999 (first entry) Chimeric signal peptide toxin gene. Toxin; androctonin; scorpion; fusion protein; transgenic plant; resistance; fungus; bacterium; infection; ss. Chimeric - Androctonus australis. OS Chimeric - Nicotiana sp. XX WO9909189-A1. FN 25-FEB-1999. PD 18-AUG-1998; 98WO-FR01814. PE 20-AUG-1997; 97FR-0010632. PR (RHON) RHONE-POULENC AGROCHIMIE. PA Derose R, Freyssinet G, Hoffmann J; PI WPI: 1999-181046/15. DR P-PADB: W. DT DNA encoding scorpion peptide androctonin - especially for producing disease-resistant plants

Claim 17, Page 26; 37pp; French.
 This sequence corresponds to the coding sequence for a fusion gene comprising the tobacco PR-1alpha gene signal peptide sequence linked to the gene encoding the toxin androctonin from the scorpion *Androctonus australis*, for expression in plants. Transgenic plants containing *androctonin* genes are stated to be resistant to fungal and bacterial infections, especially caused by *Cercospora beticola*, *Cladosporium herbarum*, *Fusarium culmorum*, *Fusarium graminearum* or *Phytophthora cinamomii*.
 Sequence 211 BP; 33 A; 58 C; 57 G; 63 T; 0 other;
 Query Match 59.6%; Score 117.4; DB 20; Length 211;
 Best Local Similarity 77.0%; Pred. No. 1.7e-26;
 Matches 161; Conservative 0; Mismatches 36; Indels 12; Gaps 1;
 QY 1 ggcgtgaagccatgggttgcgtgctttctctccaacttccacttctctgtgctca 60
 DB 1 ggcgtgaagccatgggttgcgtgctttctctccaacttccacttctctgtgctca 60
 QY 61 cctctctcttcttctctgtgactctcaactcttgcgcgtgccaagaagccagtgc 120
 DB 61 cctctctcttcttctctgtgactctcaactcttgcgcgtgccaagaagccagtgc 120
 QY 121 caatcatctactgcgaacgggaagactgtaagtgcacagaga-----tgtag 168
 DB 121 tcaagatctgcgcgaaggagggtggtgctgctactacaagltgcactaagccatactgag 180
 QY 169 ctgcgcgaaggcgaaacgtgtcgaagcagatcc 197
 DB 181 ctgcgcgaaggcgaaacgtgtcgaagcagatcc 209
 RESULT 3
 ID AAX59773 standard; cDNA: 106 BP.
 AC AAX59773;
 DT 26-JUL-1999 (first entry)
 DE cDNA encoding a thanatine fusion product of the invention.
 KW Thanatine; fungal disease; bacterial disease; *Cercospora beticola*;
 KW *Cladosporium herbarum*; *Fusarium culmorum*; *F. graminearum*;
 KW *Phytophthora cinamomii*; selection marker; plant transformation;
 KW herbicide resistance; PR-1a gene; tobacco; ss.
 OS Synthetic.
 PN FR270853-A1.
 PD 14-MAY-1999.
 PF 07-NOV-1997; 97FR-0014263.
 PR 07-NOV-1997; 97FR-0014263.
 PA (RHON) RHONE-POULENC AGROCHIMIE.
 PI Derose R, Freyssinet G, Hoffmann J;
 DR WP1; 1999-315645/27.
 DR P-PSDB; AAY15465.
 PT New nucleic acid encoding thanatine useful as a selection marker for
 PT transformation of plants
 PS Example 1; Page 14; 24pp; French.

CC The specification describes a nucleic acid sequence containing the
 CC sequence that encodes thanatine. Plants transformed with this nucleic
 CC acid sequence are resistant to fungal and bacterial diseases,
 CC specifically those caused by *Cercospora beticola*, *Cladosporium*
 CC *herbarum*, *Fusarium culmorum*, *F. graminearum* and *Phytophthora*
 CC *cinamomi*. The nucleic acid sequence may also be used as a selection
 CC marker for transformation of plants with other coding sequences,
 CC e.g. those that impart resistance to herbicides. The present sequence
 CC encodes a thanatine fusion product of the invention, comprising the
 CC signal peptide of the PR-1a gene of tobacco.
 CC
 XX Sequence 106 BP; 8 A; 34 C; 20 G; 44 T; 0 other;

Query Match 52.3%; Score 103; DB 20; Length 106;
 Best Local Similarity 100.0%; Pred. No. 3.6e-22;
 Matches 103; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggcgtcagccatgggttcgtgcttcttcacagcttcacatcttctcttgctcta 60
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1 ggcgtcagccatgggttcgtgcttcttcacagcttcacatcttctcttgctcta 60
 QY 61 ctctctcttcttcctgtgatctctcactcttcgcgtgcgg 103
 ||||||||||||||||||||||||||||||||||||||||||||||||
 Db 61 ctctctcttcttcctgtgatctctcactcttcgcgtgcgg 103

RESULT 4
 AAX29728 AAX29728 standard; DNA; 106 BP.
 XX
 AC AAX29728;

DT 22-JUN-1999 (first entry)

DE Tobacco PR-1alpha signal peptide sequence.

XX Toxin; androctonin; scorpion; fusion protein; transgenic plant;
 KM resistance; fungus; bacterium; infection; ss.

XX *Nicotiana* sp.

XX MO9909189-A1.

PD 25-FEB-1999.

PF 18-AUG-1998; 98WO-FR01814.

XX 20-AUG-1997; 97FR-0010632.

XX (RHON) RHONE-POULENC AGROCHIMIE.

PA Derose R, Freyssinet G, Hoffmann J;

PI WPI; 1999-181046/15.

DR P-PSDB; AAY99576.

XX DNA encoding scorpion peptide androctonin - especially for producing
 PT disease-resistant plants

PS Claim 15; Page 26; 37pp; French.

XX This sequence corresponds to the signal peptide sequence from the
 CC PR-1alpha gene of tobacco. The sequence is used to generate a fusion
 CC gene with the scorpion androctonin toxin gene from *Androctonus australis*,
 CC for expression in a plant. Transgenic plants containing androctonin
 CC genes are stated to be resistant to fungal and bacterial infections,
 CC especially caused by *Cercospora beticola*, *Cladosporium herbarum*, *Fusarium*
 CC *culmorum*, *Fusarium graminearum* or *Phytophthora cinamomi*.

XX Sequence 106 BP; 8 A; 34 C; 20 G; 44 T; 0 other;

Query Match 52.3%; Score 103; DB 20; Length 106;
 Best Local Similarity 100.0%; Pred. No. 3.6e-22;
 Matches 103; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggcgtcagccatgggttcgtgcttcttcacagcttcacatcttctcttgctcta 60
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1 ggcgtcagccatgggttcgtgcttcttcacagcttcacatcttctcttgctcta 60
 QY 61 ctctctcttcttcctgtgatctctcactcttcgcgtgcgg 103
 ||||||||||||||||||||||||||||||||||||||||||||||||
 Db 61 ctctctcttcttcctgtgatctctcactcttcgcgtgcgg 103

RESULT 5
 AAZ45694 AAZ45694 standard; DNA; 120 BP.
 XX
 AC AAZ45694;

DT 06-APR-2000 (first entry)

DE DNA encoding the tobacco PR-1alpha signal peptide.

XX PR-1alpha; heliomyacin; polycysteine peptide; antifungal; antibacterial;
 KM fungal infection; plant protection; animal protection; transgenic plant;
 KM fungi; bacteria; *Cercospora beticola*; *Cladosporium herbarum*;
 KM *Fusarium culmorum*; *F. oxysporum*; *Phytophthora cinamomi*; ss.

XX *Nicotiana* sp.

XX Key Location/Qualifiers

FT 12..102

FT CDS /*tag= a

FT /note= "no termination codon given"

XX FR2777568-A1.

PD 22-OCT-1999.

PF 15-APR-1998; 98FR-0004933.

XX 15-APR-1998; 98FR-0004933.

XX (RHON) RHONE-POULENC AGROCHIMIE.

PA WPI; 2000-108532/10.

DR P-PSDB; AAY54377.

XX New polycysteine peptides, designated heliomyacin, with antifungal or
 PT antibacterial activity, useful in medicine or plant protection -

PS Example 4; Page 33; 46pp; French.

XX The present sequence encodes the tobacco PR-1alpha signal peptide.
 CC This signal sequence is used to produce a fusion peptide with
 CC heliomyacin. The attachment of the PR-1alpha sequence to the N-terminal
 CC aids secretion of the heliomyacin peptide. Heliomyacin is a polycysteine
 CC peptide with antifungal and antibacterial activity, isolated from the
 CC lepidopteron *Heliothis virescens*. Heliomyacin peptides contain the
 CC sequence given in AAY54378. The heliomyacin peptides are used as
 CC pharmaceuticals for treating or preventing fungal infections in humans
 CC and animals. They may also be used as antifungal agents for plant
 CC protection. Nucleic acid encoding the heliomyacin peptides is used to
 CC generate transgenic plants that are resistant to some fungi and
 CC bacteria (specifically *Cercospora beticola*, *Cladosporium herbarum*,
 CC *Fusarium culmorum*, *F. oxysporum* and *Phytophthora cinamomi*), and for
 CC recombinant production of the peptides.

XX Sequence 120 BP; 14 A; 36 C; 23 G; 47 T; 0 other;

Query Match 51.6%; Score 101.6; DB 21; Length 120;
 Best Local Similarity 92.2%; Pred. No. 1e-21;

Db 1 ccatgggttcgtgtcttcctcagctccatcttcctctgtgtctacctcttc 60
 Oy 70 ttctctgtatctctcactctgcctgcggt 104
 Db 61 ttctctgtatctcctcactctgcctgcgcat 95

RESULT 8

AAK29730 ID AAK29730 standard; DNA; 75 BP.

AC AAK29730;

DT 22-JUN-1999 (first entry)

XX Oligo #1 for scorpion toxin fusion gene.

DE Toxin; androctonin; scorpion; fusion protein; transgenic plant;

KW resistance; fungus; bacterium; infection; ss.

OS Synthetic.

XX MO9909189-A1.

PD 25-FEB-1999.

PF 18-AUG-1998; 98WO-FR01814.

PR 20-AUG-1997; 97FR-0010632.

PA (RHON) RHONE-POULENC AGROCHIMIE.

PI Derose R, Freyssinet G, Hoffmann J;

DR WPI; 1999-181046/15.

XX DNA encoding scorpion peptide androctonin - especially for producing

PT disease-resistant plants

PS Example 1; Page 11; 37pp; French.

CC This sequence corresponds to an oligonucleotide used to generate a
 CC fusion gene (AAK29729) comprising the tobacco PR-1alpha gene signal
 CC peptide sequence linked to the gene encoding the toxin androctonin
 CC from the scorpion Androctonus australis, for expression in plants.
 CC Transgenic plants containing androctonin genes are stated to be
 CC resistant to fungal and bacterial infections, especially caused by
 CC Cercospora beticola, Cladosporium herbarum, Fusarium culmorum,
 CC Fusarium graminearum or Phytophthora cinamomi.
 SQ Sequence 75 BP; 5 A; 23 C; 13 G; 34 T; 0 other;

Query Match 37.3%; Score 73.4; DB 20; Length 75;

Best Local Similarity 98.7%; Pred. No. 3 Be-13; Indels 0; Gaps 0;

Matches 74; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 ggcgtgacgcacatgggttcgtcttcctcagctccatcttcctctgtgtcta 60

Db 1 ggcgtgacgcacatgggttcgtcttcctcagctccatcttcctctgtgtcta 60

Oy 61 cctctctctcttc 75

Db 61 cctctctctcttc 75

RESULT 9

AAK29731/C ID AAK29731 standard; DNA; 72 BP.

AC AAK29731;

DT 22-JUN-1999 (first entry)

XX Oligo #2 for scorpion toxin fusion gene.
 DE Toxin; androctonin; scorpion; fusion protein; transgenic plant;
 KW resistance; fungus; bacterium; infection; ss.

OS Synthetic.

XX MO9909189-A1.

PD 25-FEB-1999.

PF 18-AUG-1998; 98WO-FR01814.

PR 20-AUG-1997; 97FR-0010632.

PA (RHON) RHONE-POULENC AGROCHIMIE.

PI Derose R, Freyssinet G, Hoffmann J;

DR WPI; 1999-181046/15.

XX DNA encoding scorpion peptide androctonin - especially for producing
 PT disease-resistant plants

PS Example 1; Page 11; 37pp; French.

CC This sequence corresponds to an oligonucleotide used to generate a
 CC fusion gene (AAK29729) comprising the tobacco PR-1alpha gene signal
 CC peptide sequence linked to the gene encoding the toxin androctonin
 CC from the scorpion Androctonus australis, for expression in plants.
 CC Transgenic plants containing androctonin genes are stated to be
 CC resistant to fungal and bacterial infections, especially caused by
 CC Cercospora beticola, Cladosporium herbarum, Fusarium culmorum,
 CC Fusarium graminearum or Phytophthora cinamomi.
 SQ Sequence 72 BP; 33 A; 11 C; 23 G; 5 T; 0 other;

Query Match 34.2%; Score 67.4; DB 20; Length 72;

Best Local Similarity 98.6%; Pred. No. 2 Be-11; Indels 0; Gaps 0;

Matches 68; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 35 gcttcacatcttcctctgtgtcactctcttcctctgtgtgtcctcctctg 94

Db 72 gcttcacatcttcctctgtgtcactctcttcctctgtgtgtcctcctcctg 13

Oy 95 ccgtgcgcg 103

Db 12 CCGTGCCTG 4

RESULT 10

AAK59771 ID AAK59771 standard; CDNA; 63 BP.

AC AAK59771;

DT 26-JUL-1999 (first entry)

XX cDNA encoding a thanatine peptide.

DE Thanatine; fungal disease; bacterial disease; Cercospora beticola;
 KW Cladosporium herbarum; Fusarium culmorum; F. graminearum;
 KW Phytophthora cinamomi; selection marker; plant transformation;
 KW herbicide resistance; ss.

XX Psodius sp.

XX FR2770853-A1.

PD 14-MAY-1999.

```
PF 07-NOV-1997; 97FR-0014263.
XX
XX 07-NOV-1997; 97FR-0014263.
XX
PA (RHON ) RHONE-POULENC AGROCHIMIE.
XX
PI Derose R, Freyssinet G, Hoffmann J;
XX
DR WP1; 1999-315645/27.
XX
DR P-PSDB; AA1515463.
XX
PT New nucleic acid encoding thanatine useful as a selection marker for
PT transformation of plants.
XX
PS Claim 4; Page 13; 24pp; French.
XX
CC The specification describes a nucleic acid sequence containing the
CC sequence that encodes thanatine. Plants transformed with this nucleic
CC acid sequence are resistant to fungal and bacterial diseases,
CC specifically those caused by Cercospora beticola, Cladosporium
CC herbarum, Fusarium culmorum, F. graminearum and Phytophthora
CC cinamomi. The nucleic acid sequence may also be used as a selection
CC marker for transformation of plants with other coding sequences,
CC e.g. those that impart resistance to herbicides. The present sequence
CC encodes the peptide thanatine.
XX
SQ Sequence 63 BP; 19 A; 14 C; 19 G; 11 T; 0 other;

Query Match 32.08; Score 63; DB 20; Length 63;
Best Local Similarity 100.0%; Pred. No. 5,6e-10;
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 102 gttccaaagccagtcgcaatcatctactgcaacagagagctgtaagtgcagag 161
DB 1 gttccaaagccagtcgcaatcatctactgcaacagagagctgtaagtgcagag 60
OY 162 atg 164
DB 61 atg 63

RESULT 11
AAT06310
ID AAT06310 standard; DNA; 484 BP.
XX
AC AAT06310;
XX
DT 14-APR-1996 (first entry)
XX
DE Chimeric avirulence gene, Avr9.
XX
KW Pathogen resistance; Cf-9; tomato; C. fulvum; Avr 4; Avr 9; fungal;
KW leaf mould; variegation; ds.
XX
OS Lycopersicon esculentum.
OS Cladosporium fulvum.
XX
FH Key Location/Qualifiers
FT CDS 4..415
FT /tag= a
FT /product= secreted_Avr9_protein
XX
XX MO9531564-A2.
XX
XX 23-NOV-1995.
XX
XX 11-MAY-1995; 95MO-GB01075.
XX
XX 07-APR-1995; 95GB-0007232.
XX
XX 11-MAY-1994; 94GB-0009394.
XX
XX 23-DEC-1994; 94MO-GB02812.
XX
XX 31-MAR-1995; 95GB-0006658.
```

```
XX
XX (GATS-) GATSBY CHARITABLE FOUND.
XX
XX Hammond-Kosack KE, Jones DA, Jones JDG;
XX
XX WP1; 1996-010949/01.
XX
XX P-PSDB; AAR85302.
XX
PT Increasing plant pathogen resistance by induction of variegation
PT may lead to acquired resistance to a broad range of pathogens.
XX
XX
PS Claim 7; Page 93; 131pp; English.
XX
CC AAT06310 is the chimeric avirulence (Avr) gene Avr9. It is engineered
CC from the tomato signal peptide sequence of the Pr-1a gene (AAT06305) and
CC the Avr9 gene of C. fulvum. In a new method pathogen resistance genes are
CC expressed highly in genetic constructs which may be used to impart a
CC broad range of pathogen resistance, by induction of variegation, to
CC transgenic plants (or parts or propagules of plants) containing such
CC constructs. The plant pathogen resistance gene Cf-9 imparts resistance
CC to the disease caused by the leaf mould fungal pathogen Cladosporium
CC fulvum. C. fulvum contains avirulence (Avr) genes that confer recognition
CC by plants containing Cf-genes, leading to the activation of host defence
CC mechanisms to attack the disease.
XX
SQ Sequence 484 BP; 128 A; 124 C; 121 G; 111 T; 0 other;

Query Match 31.28; Score 61.4; DB 17; Length 484;
Best Local Similarity 67.7%; Pred. No. 3,4e-09;
Matches 86; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

OY 10 ccatgggttcgtgcttctctcagcttcacatcttctctctgtgtactctctc 69
DB 3 cgatggagattgtctctcttcaacaattgccttcattctctgtctctacactctct 62
OY 70 ttctccttgatctctcactcttgcgctgcggttcacaaagcagagcccaatcatt 129
DB 63 tattcctagtaataaccactcttgcgcgcgcaagcccaaaactcaacataaac 122
OY 130 actgcaa 136
DB 123 catgcaa 129

RESULT 12
AAT47876
ID AAT47876 standard; DNA; 484 BP.
XX
AC AAT47876;
XX
DT 27-MAR-1997 (first entry)
XX
DE C. fulvum avirulence AVR4 fusion gene.
XX
KW Pathogen resistance; disease resistance; AVR4 gene; avirulence;
KW tomato leaf mould; Lycopersicon; Cf-4 gene; transgenic plant;
KW crop protection; elicitor; ds.
XX
OS Chimeric Nicotiana sp.
OS Chimeric Cladosporium fulvum.
XX
FH Key Location/Qualifiers
FT CDS 5..415
FT /tag= a
FT sig_peptide 5..94
FT /tag= b
FT /note= "tobacco Pr1a protein signal sequence"
FT mat_peptide 95..412
FT /tag= c
FT /note= "C. fulvum Avr4"
XX
XX MO9635790-A1.
```

XX 14-NOV-1996.
XX 13-MAY-1996; 96MO-GB01155.
XX 11-MAY-1995; 95GB-0009575.
XX (INNE-) INNES CENT INNOVATIONS LTD JOHN.
XX Balint-Kurti PJ, Jones DA, Jones JDG, Thomas CM;
XX WPI: 1996-518678/51.
XX P-PSDB; AAM09253.
XX New plant gene Cf-4 imparting resistance to pathogens - esp.
XX Cladosporium fulvum, in tomato, and related vectors, transformed
XX cells and plants
XX Disclosure: Fig 4; 83pp; English.
XX A Clal/Sali DNA fragment (AAT47876) encodes a fusion (AAM09253) between
XX the tobacco P1a signal peptide and Cladosporium fulvum mature
XX AVR4 elicitor. The AVR4 gene, which confers avirulence to C.
XX fulvum on tomato lines carrying the pathogen resistance gene Cf-4
XX (see also AAT47877), was amplified by PCR, fused to the P1a signal
XX sequence, and the chimeric gene (SPAVR4) inserted as a Clal/Sali
XX fragment into a cDNA copy of potato virus X. The construct was
XX used in methods of assaying the function of the tomato Cf-4 gene
XX in plants and of identifying individuals potentially carrying the
XX Cf-4 gene.
SQ Sequence 484 BP; 128 A; 124 C; 121 G; 111 T; 0 other;
Query Match 31.2%; Score 61.4; DB 17; Length 484;
Best Local Similarity 67.7%; Pred. No. 3.4e-09;
Matches 86; Conservative 0; Mismatches 41; Indels 0; Gaps 0;
QY 10 ccatggttcgtctctctcagctccatcttctctctgtgtactctctc 69
DB 3 ccatggttcgtctctctcagctccatcttctctctgtgtactctctc 62
QY 70 tttctctgtgactctcactctctgcgcgtgcaggtccaaagcagccatcatc 129
DB 63 tattcctagtaataaccactctctgcgtgcagcccaaacactcaacatacaacc 122
QY 130 actgcaa 136
DB 123 catgcaa 129
RESULT 13
AADI0075
ID AADI0075 standard; DNA; 107 BP.
XX AADI0075;
XX 18-SEP-2001 (first entry)
XX P1a leader sequence DNA.
XX Heavy chain immunoglobulin; pathogen resistance; metabolism modulator;
XX passive immunisation; P1a leader peptide; ds.
XX unidentified.
XX OS
XX Key Location/Qualifiers
FH CDS 3..104
FT /tag= a
FT /product= "P1a leader peptide"
FT /note= "CDS does not include stop codon"
FT /partial
XX

PN EP118669-A2.
XX 25-JUL-2001.
XX 08-DEC-2000; 2000EP-0310997.
XX 17-DEC-1999; 99EP-0310188.
XX (UNIL) UNILEVER PLC.
XX (UNIL) UNILEVER NV.
XX Frenken LGJ, Van Der Logt CPE, Jobling SA, Teh Y;
XX WPI: 2001-427157/46.
XX P-PSDB; AAE05290.
XX Modifying a plant to produce an antibody useful for increasing pathogen
XX resistance or to modulate metabolism comprises introducing a DNA
XX sequence encoding a heavy chain immunoglobulin linked to a peptide that
XX targets a cellular compartment -
XX Example 1; Page 24; 81pp; English.
XX The present invention relates to a method for modifying a plant to
XX produce an antibody or an active fragment or derivative, or a protein
XX functional equivalent, in a cellular compartment comprising introducing
XX a DNA sequence encoding a heavy chain immunoglobulin, where the DNA is
XX linked to promoters and provided with an additional sequence encoding a
XX peptide capable of targeting heavy chain immunoglobulin to a cellular
XX compartment. The method is used for producing a heavy chain
XX immunoglobulin or an active fragment or derivative, or a protein that is
XX functionally equivalent for increasing the pathogen resistance in a plant
XX or to modulate metabolism in a plant. Under some circumstances it may be
XX desirable to retain the antibody product with the plant rather than
XX extracting and isolating the product. In particular, edible selected
XX antigens may be used in a method of passively immunising an animal,
XX preferably human, against the antigen, e.g., pathogenic organisms. The
XX present sequence is a P1a leader sequence DNA which is used in the
XX exemplification of the invention.
SQ Sequence 107 BP; 17 A; 32 C; 16 G; 42 T; 0 other;
Query Match 29.6%; Score 58.4; DB 22; Length 107;
Best Local Similarity 77.2%; Pred. No. 1.7e-08;
Matches 71; Conservative 0; Mismatches 21; Indels 0; Gaps 0;
QY 10 ccatggttcgtctctcagctccatcttctctctgtgtactctctc 69
DB 1 ccatggttcgtctctcagctccatcttctctctgtgtactctctc 60
QY 70 tttctctgtgactctcactctctgcgcgtgcaggtccaaagcagccatcatc 101
DB 61 tattcctagtaataaccactctctgcgtgcagcccaaacactcaacatacaacc 92
RESULT 14
AAQ06180
ID AAQ06180 standard; DNA; 809 BP.
XX AAQ06180;
XX 31-JAN-1991 (first entry)
XX PR-1a cDNA cloned into plasmid pBSpr1-207.
XX Transgenic plants; disease resistance; chimeric DNA;
XX plant pathogenesis-related protein; ss.
XX synthetic.
XX OS
XX Key Location/Qualifiers
FH mat_peptide 123..0
FT

```
FT      /*tag- a
FT      /product=PR-1a
XX      EP992225-A.
XX      17-OCT-1990.
XX      21-MAR-1990; 90EP-0105336.
XX      20-OCT-1989; 89US-0425504.
XX      24-MAR-1989; 89US-0329018.
XX      20-JUN-1989; 89US-0368672.
XX      (CIBA ) CIBA GEIGY AG.
XX      Ryals JA, Alexander DC, Goodman RM, Melns F, Payne GB;
XX      Stinson JR, Neuhaus J-M, Moyer MB;
XX      WPI: 1990-313983/42.
XX      P-PSDB: AAR07313.
XX      Disease-resistant transgenic plants - obt'd. using encoding an
XX      inducible pathogenesis-related protein from infected plants.
XX      Example 15; page 23; 77pp: English.
XX      This chimeric DNA sequence comprises the PR-1a plant pathogenesis-
XX      related protein (PRP)-coding sequence (I), and a promoter sequence
XX      which enhances transcription of (I). This construct is used to
XX      produce transgenic plant cells or -tissues with the ability to re-
XX      generate into plants which are disease resistant.
XX      See also AA006179, AA006181-86, AA006199-Q06208 and AA006829.
XX      Sequence 809 BP; 231 A; 153 C; 171 G; 254 T; 0 other:

Query Match      29.3%; Score 57.8; DB 11; Length 809;
Best Local Similarity 76.3%; Pred. No. 5.2e-08;
Matches 71; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY      9 gccatggattcgctctctcagctccatctctcctctctgtgctactctctt 68
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      27 gtcattggattgtctctcttcacaaatgccttcattctctctgtctcaactctc 86

QY      69 ctttcctgtgactctcaactctgtccgtgcc 101
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      87 ttattcctagtaataatacccaactctgtccgtgcc 119

RESULT 15
AAV62305
ID      AAV62805 standard; CDNA; 809 BP.
XX
AC      AAV62805;
XX
DT      05-MAR-1999 (first entry)
XX
DE      Tobacco PR-1a gene clone.
XX
KW      Chemically regulatable DNA promoter: expression control; pesticide;
XX      herbicide tolerance; pathogenesis related gene; PR gene; ss.
XX      Nicotiana acuminata.
XX      US5851766-A.
XX      22-DEC-1998.
XX      31-MAY-1995; 95US-0456262.
XX      31-MAY-1995; 95US-0456262.
XX      (NOVS ) NOVARTIS FINANCE CORP.
```

```
XX      Harms C, Ryals JA;
XX      WPI: 1999-080396/07.
XX      Isolating chemically regulatable DNA sequences in plants - useful
XX      for chemically controlling expression in transformed plants
XX      Example 44; Column 179-182; 175pp: English.
XX      This sequence represents a clone of the tobacco pathogenesis related (PR)
XX      gene. This gene can be isolated using the method of the invention.
XX      The method is for isolating a chemically regulatable DNA promoter
XX      fragment from the 5' flanking region of a chemically regulatable gene in
XX      a plant tissue. The method allows isolation of sequences which will be
XX      useful for the controlled expression of genes, under the control of a
XX      non-coding regulatable sequence. This is useful in plants with a
XX      herbicide or pesticide detoxification mechanism under the control of a
XX      chemical regulator, the regulator being applied before or with the
XX      herbicide or pesticide to give optimal tolerance. The promoter fragment
XX      is useful for controlling sequences which encode traits such as
XX      height, shape, development, male or female sterility, and the ability
XX      of the plant to withstand cold, heat, salt and drought. The chemical
XX      induction of the promoter allows the regulation of production of
XX      compounds, e.g. flavours, fragrances, pigments, natural sweeteners,
XX      industrial feedstocks, antimicrobials and pharmaceuticals, by
XX      biosynthesis or metabolic conversion, whose biosynthesis is controlled
XX      by endogenous or foreign genes. The method allows control over the time
XX      and rate of gene expression either throughout the whole plant, or in
XX      localized tissues, to achieve e.g. fungal or insect resistance by for
XX      instance dusting the leaves with the chemical regulator. Controlling the
XX      developmental processes by the application of a regulating chemical in
XX      e.g. the commercial production of cultivated crops allows processes such
XX      as germination, flower formation and fruit ripening to be synchronised at
XX      a given time.
XX      Sequence 809 BP; 231 A; 153 C; 171 G; 254 T; 0 other:

Query Match      29.3%; Score 57.8; DB 20; Length 809;
Best Local Similarity 76.3%; Pred. No. 5.2e-08;
Matches 71; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY      9 gccatggattcgctctctcagctccatctctctctctgtgctactctctt 68
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      27 gtcattggattgtctctcttcacaaatgccttcattctctctgtctcaactctc 86

QY      69 ctttcctgtgactctcaactctgtccgtgcc 101
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      87 ttattcctagtaataatacccaactctgtccgtgcc 119
```

Search completed: May 12, 2002, 20:20:40
Job time: 4927 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 12, 2002, 20:14:35 ; Search time 76.1 Seconds
(without alignments)
203.350 Million cell updates/sec

Title: US-09-554-024-2
Perfect score: 63
Sequence: 1 ggtccagaagccagtcgccc.....ctgtaagtcgacagagatg 63

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_NA:*
1: /cgn2_6/prodata/2/ina/5A.COMB.seq:*
2: /cgn2_6/prodata/2/ina/5B.COMB.seq:*
3: /cgn2_6/prodata/2/ina/5A.COMB.seq:*
4: /cgn2_6/prodata/2/ina/5B.COMB.seq:*
5: /cgn2_6/prodata/2/ina/5A.COMB.seq:*
6: /cgn2_6/prodata/2/ina/5B.COMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	23.4	37.1	3102	1	US-08-462-484-5
2	23.4	37.1	3102	1	US-08-441-147-5
3	23.4	37.1	3102	5	PCT-US95-07536-5
4	22.8	36.2	80	1	US-08-418-859-54
5	22.8	36.2	80	1	US-08-418-859-56
6	22.8	36.2	80	2	US-08-643-181-54
7	22.8	36.2	80	2	US-08-643-181-56
8	22.8	36.2	80	2	US-08-643-181-54
9	21.8	34.6	180	5	PCT-US93-03076-1
10	21.6	34.3	330	5	US-08-358-160-145
11	21.6	34.3	375	5	PCT-US91-02766-13
12	21.6	34.3	395	6	PCT-US91-02766-15
13	21.6	34.3	376	6	5428135-3
14	21.6	34.3	376	6	US-08-998-416-868
15	21.6	34.3	649	6	5194596-12
16	21.6	34.3	748	6	5219739-12
17	21.6	34.3	748	1	US-08-387-845-1
18	21.6	34.3	748	2	US-08-778-275-1
19	21.6	34.3	1247	5	US-08-867-352-1
20	21.6	34.3	1247	5	PCT-US91-02766-19
21	21.6	34.3	1247	5	5219759-3
22	21.6	34.3	1316	5	PCT-US91-02766-21
23	21.2	33.7	80	1	US-08-418-859-55
24	21.2	33.7	80	2	US-08-643-181-55
25	21.2	33.7	1074	2	US-08-627-151A-15
26	21.2	33.7	1404	6	5171840-8
27	21.2	33.7	1486	4	5480796-8

28	21.2	33.7	2061	6	5171840-1	Patent No. 5171840
29	21.2	33.7	2061	6	5480796-1	Patent No. 5480796
30	21.2	33.7	3319	4	US-08-795-473B-2	Sequence 2, Appl
31	21.2	33.7	3319	4	US-08-358-160-143	Sequence 143, App
32	21.2	33.3	234	1	US-08-253-155A-8	Sequence 8, Appl
33	21.2	33.3	658	3	US-08-705-771-11	Sequence 11, Appl
34	21.2	33.3	658	3	US-09-109-204-27	Sequence 27, Appl
35	21.2	33.3	1882	3	US-08-501-572-4	Sequence 4, Appl
36	21.2	33.3	1882	3	US-09-040-444-4	Sequence 4, Appl
37	21.2	33.3	2224	3	US-09-109-204-6	Sequence 6, Appl
38	20.8	33.0	1773	4	US-08-943-731-215	Sequence 215, App
39	20.8	33.0	1966	6	5475086-3	Patent No. 5475086
40	20.8	33.0	2026	1	US-08-485-718-8	Sequence 8, Appl
41	20.8	33.0	2026	2	US-08-484-530-54	Sequence 54, Appl
42	20.8	33.0	2026	2	US-08-827-618A-54	Sequence 54, Appl
43	20.8	33.0	2026	3	US-08-483-952A-54	Sequence 54, Appl
44	20.8	33.0	6246	4	US-08-943-731-640	Sequence 640, App
45	20.6	32.7	699	4	US-08-991-789A-172	Sequence 172, App

ALIGNMENTS

RESULT 1
US-08-462-484-5
Sequence 5, Application US/08462484
Patent No. 5667331
GENERAL INFORMATION:
APPLICANT: Yaver, Debbie Sue
APPLICANT: Xu, Feng
APPLICANT: Dalboge, Henrik
APPLICANT: Schneider, Palle
APPLICANT: Aslyng, Dorrit A.
TITLE OF INVENTION: PURIFIED POLYPORUS LACCASES AND
TITLE OF INVENTION: NUCLEIC ACIDS ENCODING SAME
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESSES:
ADDRESS: No. 56675310 No. 5667531disk of No. 5667531th America, Inc.
STREET: 405 Lexington Avenue, Suite 6400
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/462,484
FILING DATE: 05-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/441,147
FILING DATE: 15-MAY-1995
ATTORNEY/AGENT INFORMATION:
NAME: Lowmey, Karen A.
REGISTRATION NUMBER: 31,274
REFERENCE/DOCKET NUMBER: 4185.010-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 867 0123
TELEFAX: 212 878 9655
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 3102 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Polyporus plinthus
FEATURE:
NAME/KEY: intron

[illegible]

Query Match	37.1%	Score	23.4	DB	1	Length	3102
Best Local Similarity	73.2%	Pred. No.	6.5				
Matches	30	Mismatches	11	Indels	0	Gaps	0
Conservative							

QY	14	cagttgccaatcatctactctgcaacagagagactgtgtaagtgc	54
Db	2614	CACAGTGACTCATCTTTTGGCAACAGGAGACTGGACAACGC	2654

RESULT 2
US-08-441-147-5
Sequence 5, Application US/08441147
Patent NO. 5770418
GENERAL INFORMATION:
APPLICANT: Yaver, Debbie Sue
APPLICANT: Xu, Feng
APPLICANT: Dalboge, Henrik
APPLICANT: Schneider, Palle
APPLICANT: Aaslyng, Dorrit A.
TITLE OF INVENTION: PURIFIED POLYPORUS LACCASES AND
TITLE OF INVENTION: NUCLEIC ACIDS ENCODING SAME
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESSES:
ADDRESSEE: No. 57704180 No. 5770418disk of No. 5770418th America, Inc
STREET: 405 Lexington Avenue, Suite 6400
City: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/441, 147
FILING DATE: 15-MAY-1995

```

1 CLASSIFICATION: 435
2
3 ATTORNEY/AGENT INFORMATION:
4   NAME: Lowney, Karen A.
5   REGISTRATION NUMBER: 31,274
6   REFERENCE/DOCKET NUMBER: 4185, 010-US
7   TELECOMMUNICATION INFORMATION:
8     TELEPHONE: 212 867 0123
9     TELEFAX: 212 878 9655
10    INFORMATION FOR SEQ ID NO: 5:
11      SEQUENCE CHARACTERISTICS:
12        LENGTH: 3102 base pairs
13        TYPE: nucleic acid
14        STRANDEDNESS: double
15        TOPOLOGY: linear
16      MOLECULE TYPE: DNA (genomic)
17      ORIGINAL SOURCE:
18        ORGANISM: Polyporus plustus
19
20      FEATURE:
21        NAME/KEY: intron
22        LOCATION: 666..720
23
24      FEATURE:
25        NAME/KEY: intron
26        LOCATION: 790..845
27
28      FEATURE:
29        NAME/KEY: intron
30        LOCATION: 1125..1182
31
32      FEATURE:
33        NAME/KEY: intron
34        LOCATION: 1390..1450
35
36      FEATURE:
37        NAME/KEY: intron
38        LOCATION: 1607..1661
39
40      FEATURE:
41        NAME/KEY: intron
42        LOCATION: 1863..1918
43
44      FEATURE:
45        NAME/KEY: intron
46        LOCATION: 1976..2025
47
48      FEATURE:
49        NAME/KEY: intron
50        LOCATION: 2227..2285
51
52      FEATURE:
53        NAME/KEY: intron
54        LOCATION: 2403..2458
55
56      FEATURE:
57        NAME/KEY: intron
58        LOCATION: 2576..2627
59
60      NAME/KEY: CDS
61      LOCATION: 1665..721, 789..846, 1124..1183, 1389..1451,
62      1666..1662, 1862..1919, 1975..2026, 2226..2286, 2402..2459,
63      2575..2628).
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
```

```

Query March 37.1%; Score 23.4; Db 1, Length 3102;
Best Local Similarity 73.2%; Pred. No. 6.5;
Matches 30; Conservative 0; Mismatches 11; Indels 0; Gaps 0.

Oy 14 cagtcgcaatcatctactgcgaacagagagactgtaagtgc 54
      ||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2614 CACAGTGACTCATCTTTTGCAACAGGAGGACHTGGAACAAGC 2654

RESULT 3
PCT-US95-07536-5
: Sequence 5, Application PC/TUS9507536
: GENERAL INFORMATION:
: APPLICANT:
: TITLE OF INVENTION:
: TITLE OF INVENTION: PURIFIED POLYPORUS LACCASES AND
: NUCLEIC ACIDS ENCODING SAME
: NUMBER OF SEQUENCES: 10

```

CORRESPONDENCE ADDRESS:
ADDRESSEE: Novo Nordisk of North America, Inc.
STREET: 405 Lexington Avenue, Suite 6400
COUNTRY: U.S.A.
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/07536
FILING DATE: 15-June-1995
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 08/265,534
FILING DATE: 24-June-1994
ATTORNEY/AGENT INFORMATION:
NAME: Lowmey, Karen A.
REGISTRATION NUMBER: 31,274
REFERENCE/DOCKET NUMBER: 4185.204-WO
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 867 0123
TELEFAX: 212 878 9635
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 3102 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORGANISM: Polyporus pinastus
FEATURE:
NAME/KEY: intron
LOCATION: 666..720
FEATURE:
NAME/KEY: intron
LOCATION: 790..845
FEATURE:
NAME/KEY: intron
LOCATION: 1125..1182
FEATURE:
NAME/KEY: intron
LOCATION: 1390..1450
FEATURE:
NAME/KEY: intron
LOCATION: 1607..1661
FEATURE:
NAME/KEY: intron
LOCATION: 1863..1918
FEATURE:
NAME/KEY: intron
LOCATION: 1976..2025
FEATURE:
NAME/KEY: intron
LOCATION: 2227..2285
FEATURE:
NAME/KEY: intron
LOCATION: 2403..2458
FEATURE:
NAME/KEY: intron
LOCATION: 2576..2627
FEATURE:
NAME/KEY: CDS
LOCATION: join (665..721, 789..846, 1124..1183, 1389..1451, 1606..1662, 1862..1919, 1975..2026, 2226..2286, 2402..2459, 2575..2628).
PCT-US95-07536-5

Query Match 37.1%; Score 23.4; DB 5; Length 3102;
Best Local Similarity 73.2%; Pred. No. 6.5; 11; Indels 0; Gaps 0;
Matches 30; Conservative 0; Mismatches 11;

QY 14 cagtcgcaatcatctactgcaacagagagcctgtaagtc 54
Db 2614 CACAGTGCATCATCTTTTGCACAGAGAGCTGACACAGC 2654

RESULT 4
US-08-418-859-54
Sequence 54, Application US/08418859
Patent No. 5811235
GENERAL INFORMATION:
APPLICANT: Jeffreys, Alec J.
TITLE OF INVENTION: METHOD OF
TITLE OF INVENTION: CHARACTERISATION
NUMBER OF SEQUENCES: 57
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cushman, Darby & Cushman
STREET: 1100 New York Avenue, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005-3918
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.2 Mb
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS 3.20
SOFTWARE: ASCII from WPS-PLUS
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/418,859
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/935,107
FILING DATE: 26 August 1992
APPLICATION NUMBER: 9118371.5
FILING DATE: 27-Aug-1991
APPLICATION NUMBER: 9119089.2
FILING DATE: 06-Sep-1991
APPLICATION NUMBER: 9124636.3
FILING DATE: 20-No. 5811235-1991
APPLICATION NUMBER: 9207379.0
FILING DATE: 03-Apr-1992
APPLICATION NUMBER: 9212627.5
FILING DATE: 15-Jun-1992
APPLICATION NUMBER: 9212861.8
FILING DATE: 17-Jun-1992
ATTORNEY/AGENT INFORMATION:
NAME: KOKULIS, PAUL N.
REGISTRATION NUMBER: 16,773
REFERENCE/DOCKET NUMBER: 97279/PHM.36520/US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (292) 861-3000
TELEFAX: (292) 822-0944
INFORMATION FOR SEQ ID NO: 54:
SEQUENCE CHARACTERISTICS:
LENGTH: 80 base pairs
TYPE: Nucleic Acid
STRANDEDNESS: Single
TOPOLOGY: Linear
US-08-418-859-54

Query Match 36.2%; Score 22.8; DB 1; Length 80;
Best Local Similarity 66.0%; Pred. No. 3.1; 17; Indels 0; Gaps 0;
Matches 33; Conservative 0; Mismatches 17;

QY 10 aagccagtgccaatcatctactgcaacagagagcctgtaagtcagag 59
Db 26 AACCTAATTCGATTGGCTACTTAAAGAGAGAGGGGTATGAGCCAGAG 75

RESULT 5
US-08-418-859-56

```
; Sequence 56, Application US/08418859
; Patent No. 5811235
; GENERAL INFORMATION:
; APPLICANT: Jeffreys, Alec J.
; TITLE OF INVENTION: METHOD OF
; TITLE OF INVENTION: CHARACTERISATION
; NUMBER OF SEQUENCES: 57
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Cushman, Darby & Cushman
; STREET: 1100 New York Avenue, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3918
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.2 Mb
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS 3.20
; SOFTWARE: ASCII from WPS-PLUS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/418, 859
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/935,107
; FILING DATE: 26 August 1992
; APPLICATION NUMBER: 9118371.5
; FILING DATE: 27-Aug-1991
; APPLICATION NUMBER: 9119089.2
; FILING DATE: 06-Sep-1991
; APPLICATION NUMBER: 9124636.3
; FILING DATE: 20-No. 5811235-1991
; APPLICATION NUMBER: 9207379.0
; FILING DATE: 03-Apr-1992
; APPLICATION NUMBER: 9212627.5
; FILING DATE: 15-Jun-1992
; APPLICATION NUMBER: 9212881.8
; FILING DATE: 17-Jun-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: KOKULIS, PAUL N.
; REGISTRATION NUMBER: 16,773
; REFERENCE/DOCKET NUMBER: 97279/PHM.36520/US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (292) 861-3000
; TELEFAX: (202) 822-0944
; INFORMATION FOR SEQ ID NO: 56:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 80 Base Pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; US-08-418-859-56

Query Match          36.2%; Score 22.8; DB 1; Length 80;
Best Local Similarity 66.0%; Pred. No. 3.1;
Matches 33; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 10 aagccagtgccatcatctactgcacagagactgtaagtccagag 59
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 26 AAGCTAATTCGATTGGCTACTTTAAAGAGACAGGCGTATGACGACAG 75

RESULT 6
US-08-643-181-54
; Sequence 54, Application US/08643181
; Patent No. 5853989
; GENERAL INFORMATION:
; APPLICANT: Jeffreys, Alec J.
; TITLE OF INVENTION: METHOD OF
; TITLE OF INVENTION: CHARACTERISATION
; NUMBER OF SEQUENCES: 57
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Cushman, Darby & Cushman
; STREET: 1100 New York Avenue, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
```

```
; ADDRESSEE: Cushman, Darby & Cushman
; STREET: 1100 New York Avenue, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3918
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.2 Mb
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS 3.20
; SOFTWARE: ASCII from WPS-PLUS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/643,181
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/418,859
; FILING DATE:
; APPLICATION NUMBER: 07/935,107
; FILING DATE: 26 August 1992
; APPLICATION NUMBER: 9118371.5
; FILING DATE: 27-Aug-1991
; APPLICATION NUMBER: 9119089.2
; FILING DATE: 06-Sep-1991
; APPLICATION NUMBER: 9124636.3
; FILING DATE: 20-No. 5853989-1991
; APPLICATION NUMBER: 9207379.0
; FILING DATE: 03-Apr-1992
; APPLICATION NUMBER: 9212627.5
; FILING DATE: 15-Jun-1992
; APPLICATION NUMBER: 9212881.8
; FILING DATE: 17-Jun-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: KOKULIS, PAUL N.
; REGISTRATION NUMBER: 16,773
; REFERENCE/DOCKET NUMBER: 97279/PHM.36520/US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (292) 861-3000
; TELEFAX: (202) 822-0944
; INFORMATION FOR SEQ ID NO: 54:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 80 Base Pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; US-08-643-181-54

Query Match          36.2%; Score 22.8; DB 2; Length 80;
Best Local Similarity 66.0%; Pred. No. 3.1;
Matches 33; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 10 aagccagtgccatcatctactgcacagagactgtaagtccagag 59
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 26 AAGCTAATTCGATTGGCTACTTTAAAGAGACAGGCGTATGACGACAG 75

RESULT 7
US-08-643-181-56
; Sequence 56, Application US/08643181
; Patent No. 5853989
; GENERAL INFORMATION:
; APPLICANT: Jeffreys, Alec J.
; TITLE OF INVENTION: METHOD OF
; TITLE OF INVENTION: CHARACTERISATION
; NUMBER OF SEQUENCES: 57
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Cushman, Darby & Cushman
; STREET: 1100 New York Avenue, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
```



```

1      ZIP: 20005--3918
2      COMPUTER READABLE FORM:
3      MEDIUM TYPE: Diskette, 3.50 inch, 1.2 Mb
4      MEDIUM TYPE: storage
5      COMPUTER: IBM PS/2
6      OPERATING SYSTEM: PC-DOS 3.20
7      SOFTWARE: ASCII from WPS-PLUS
8      CURRENT APPLICATION DATA:
9      APPLICATION NUMBER: US/08/643,181
10     FILING DATE:
11     CLASSIFICATION:
12     PRIOR APPLICATION DATA:
13     APPLICATION NUMBER: US/08/418,859
14     FILING DATE:
15     APPLICATION NUMBER: 07/935,107
16     FILING DATE: 26 August 1992
17     APPLICATION NUMBER: 9118371.5
18     FILING DATE: 27-Aug-1991
19     APPLICATION NUMBER: 9119089.2
20     FILING DATE: 06-Sep-1991
21     APPLICATION NUMBER: 9124636.3
22     FILING DATE: 20-No. 5853989-1991
23     APPLICATION NUMBER: 9207379.0
24     FILING DATE: 03-Apr-1992
25     APPLICATION NUMBER: 921627.5
26     FILING DATE: 15-Jun-1992
27     APPLICATION NUMBER: 9212881.8
28     FILING DATE: 17-Jun-1992
29     ATTORNEY/AGENT INFORMATION:
30     NAME: KOKULIS, PAUL N.
31     REGISTRATION NUMBER: 16,773
32     REFERENCE/DOCKET NUMBER: 97279/PHW.36520/US
33     TELECOMMUNICATION INFORMATION:
34     TELEPHONE: (292) 861-3000
35     TELEFAX: (202) 822-0944
36     INFORMATION FOR SEQ. ID NO: 56:
37     SEQUENCE CHARACTERISTICS:
38     LENGTH: 80 Base Pairs
39     TYPE: Nucleic Acid
40     STRANDEDNESS: Single
41     TOPOLOGY: Linear
42     US-08-643-181-56
43
44     Query Match          36.2%; Score 22.8; DB 2; Length 80;
45     Best Local Similarity 66.0%; Pred. No. 3.1;
46     Matches 33; Conservative 0; Mismatches 17; Indels 0; Gaps 0.
47
48     Qy      10 aagcagctgcacatcatctcactgcacaaaggaagactgtaagtccagag 59
49             |||| | | | | | | | | | | | | | | | | | | | | | | | |
50     Db      26 AAGCTAATTCGAGTTGGCTACTTTAAAGAGACAGCAGGSGTATGACCCAGAG 75
51
52 RESULT      8
53 PCT-US93-03076-1/c
54     Sequence 1, Application PC/TUS9303076
55     GENERAL INFORMATION:
56     APPLICANT: Whitehead Institute for Biomedical Research
57     TITLE OF INVENTION: GAP-Associated Protein p190 and
58     TITLE OF INVENTION: Transduction
59     NUMBER OF SEQUENCES: 20
60     CORRESPONDENCE ADDRESS:
61     ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
62     STREET: 2 Millitia Drive
63     CITY: Lexington
64     STATE: MA
65     COUNTRY: US
66     ZIP: 02173
67     COMPUTER READABLE FORM:
68     MEDIUM TYPE: Floppy disk
69     COMPUTER: IBM PC compatible
70     OPERATING SYSTEM: PC-DOS/MS-DOS
71     SOFTWARE: PatentIn Release #1.0, Version #1.25

```

```

CURRENT APPLICATION DATA:      ? CURRENT APPLICATION NUMBER: PCT-US93-/03076
APPLICATION NUMBER:             ? FILING DATE:    19930331
CLASSIFICATION:                ? ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia       ? REGISTRATION NUMBER:   32,227
REFERENCE/DOCKET NUMBER: WHI92-03A
TELECOMMUNICATION INFORMATION: ? TELEPHONE: 617-861-6240
                                ? TELEFAX: 617-861-9540
INFORMATION FOR SEQ ID NO:     ? SEQUENCE CHARACTERISTICS:
LENGTH: 8298 base pairs        ? TYPE: NUCLEIC ACID
STRANDEDNESS: double           ? TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)   ? FEATURE:
NAME/KEY: CDS                  ? LOCATION: 731..5272
PCT-US93-03076-1               ?

Query Match          36.2% Score 22.8 DB 5 Length 8298:
Best Local Similarity 62.1%; Fred No. 16;
Matches 36; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY         1 gttcccaagagcagtgcgaatctacttcgacaggaggattgttaagtggcgaga 58
            | ||||| || | | | | | | | | | | | | | | | | | | | | | | | |
Db  4761 GCATCCACCAAGCATGTGCGTGCATTGCTGTATGTACCAGAAGGGGCTGTTACTTCGAGA 4704


RESULT      9
US-08-358-160-145
Sequence 145, Application US/08358160
Patent No. 5663143
GENERAL INFORMATION:
APPLICANT: LEY, Arthur C.
APPLICANT: LADNER, Robert C.
APPLICANT: GUTERMAN, Sonia K.
APPLICANT: ROBERTS, Bruce L.
APPLICANT: MARKLAND, William
APPLICANT: KENT, Rachel B.
TITLE OF INVENTION: ENGINEERED HUMAN-DERIVED KUNITZ
NUMBER OF SEQUENCES: 234
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: District of Columbia
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/Ms-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/358,160
FILING DATE: 16-Dec-1994
CLASSIFICATION: 514
PRIOR APPLICATION NUMBER: US 08/133,031
FILING DATE: 13-OCT-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/009,319
FILING DATE: 26-JAN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/664,989
FILING DATE: 01-MAR-1991
PRIOR APPLICATION DATA:
```

APPLICATION NUMBER: US 07/487,063
FILING DATE: 02-MAR-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/240,160
FILING DATE: 02-SEP-1988
ATTORNEY/AGENT INFORMATION:
NAME: Cooper, Iver P.
REGISTRATION NUMBER: 28,005
REFERENCE/DOCKET NUMBER: LEY-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
TELEX: 248633
INFORMATION FOR SEQ ID NO: 145:
SEQUENCE CHARACTERISTICS:
LENGTH: 180 bases
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: synthetic DNA fragment
US-08-358-160-145

Query Match 34.6%; Score 21.8; DB 1; Length 180;
Best Local Similarity 69.7%; Pred. No. 10;
Matches 23; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 26 tctactgcaacagagagactgtaagtcacaga 58
:||||| ||||| ||| ||||| ||||| |||||
Db 62 DSTACTACACTCGAGTATGCTGAAGTCSDGV 94

RESULT 10
PCT-US91-02766-13
Sequence 13, Application PC/TUS9102766
GENERAL INFORMATION:
APPLICANT: NASCIMENTO, CARLOS G.
APPLICANT: CALDERON-CACIA, MARIA D.
TITLE OF INVENTION: GLYCOSYLATED PDGF
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: Irell & Manella
STREET: 545 Middlefield Road, Suite 200
CITY: Menlo Park
STATE: California
COUNTRY: USA
ZIP: 94025
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US91/02766
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/515,474
FILING DATE: 26-APR-1990
ATTORNEY/AGENT INFORMATION:
NAME: ROBINS, ROBERTA L.
REGISTRATION NUMBER: 33,208
REFERENCE/DOCKET NUMBER: 2300-0105.40
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 327-7250
TELEFAX: (415) 327-2951
TELEX: 706141
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 330 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single

TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 1..330
PCT-US91-02766-13

Query Match 34.3%; Score 21.6; DB 5; Length 330;
Best Local Similarity 75.0%; Pred. No. 15;
Matches 27; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 22 atcatctactgcaacagagagactgtaagtcacag 57
| | | | | | | | | | | | | | | | | | | | | |
Db 130 ACCGGCTGCTGCAACACGACGAGTGTCAAGTCACG 165

RESULT 11
PCT-US91-02766-15
Sequence 15, Application PC/TUS9102766
GENERAL INFORMATION:
APPLICANT: NASCIMENTO, CARLOS G.
APPLICANT: CALDERON-CACIA, MARIA D.
TITLE OF INVENTION: GLYCOSYLATED PDGF
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: Irell & Manella
STREET: 545 Middlefield Road, Suite 200
CITY: Menlo Park
STATE: California
COUNTRY: USA
ZIP: 94025
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US91/02766
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/515,474
FILING DATE: 26-APR-1990
ATTORNEY/AGENT INFORMATION:
NAME: ROBINS, ROBERTA L.
REGISTRATION NUMBER: 33,208
REFERENCE/DOCKET NUMBER: 2300-0105.40
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 327-7250
TELEFAX: (415) 327-2951
TELEX: 706141
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 375 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 1..375
PCT-US91-02766-15

Query Match 34.3%; Score 21.6; DB 5; Length 375;
Best Local Similarity 75.0%; Pred. No. 16;
Matches 27; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 22 atcatctactgcaacagagagactgtaagtcacag 57
| | | | | | | | | | | | | | | | | | | | | |
Db 130 ACCGGCTGCTGCAACACGACGAGTGTCAAGTCACG 165

RESULT 12

Matches	27;	Conservative	0;	Mismatches	9;	Indels	0;	Gaps	0;
Oy	22	atcatctactgcgaacagaggagactggttaagtgccag		57					
Db	396	accggtctgtgcacacgcagcagtgctcagtgccag		431					

Search completed: May 12, 2002, 20:14:36
 Job time: 7193 sec

Gencore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 12, 2002, 20:20:36 ; Search time 333.08 Seconds

(without alignments)
324.744 Million cell updates/sec

Title: US-09-554-024-2

Sequence: 1 ggtccagaagaacagtgcc.....ctgtatgagcagagatg 63

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N.GeneSeq_032802.*
1: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	63	100.0	63	20	AAK59771
2	63	100.0	98	20	AAK59772
3	63	100.0	197	20	AAK59774
4	33	52.4	33	20	AAK59770
5	26.8	42.5	589	22	AAH87754
6	26.2	41.6	2022	21	AAH12996
7	25	39.7	1710	21	AAH07491
8	24.6	39.0	1599	22	AAH91430
9	24.6	39.0	1599	23	AB198000

C	10	24.6	39.0	2253	10	AAH92071	Human muscarinic a
C	11	24.6	39.0	2261	22	AAH91429	Human CHRM5 gene s
C	12	24.4	38.7	17745	23	AAH92597	DNA encoding novel
C	13	24.2	38.4	2604	24	ABH05165	Human gamma glutam
C	14	24	38.1	31122	14	AAH040706	Bacillus subtilis
C	15	23.6	37.5	4041	23	ABH18273	Drosophila melanog
C	16	23.6	37.5	6369	23	ABH18272	Drosophila melanog
C	17	23.6	37.5	13497	23	ABH03557	Drosophila melanog
C	18	23.6	37.5	16489	23	ABH03556	Drosophila melanog
C	19	23.4	37.1	439	21	AAH15619	Human secreted pro
C	20	23.4	37.1	3102	17	AAH15600	Laccase-IC33 gene.
C	21	23	36.5	459	22	AAH88664	Human digestive sy
C	22	23	36.5	781	22	AAH08150	Human CDNA clone (
C	23	23	36.5	803	22	AAH03502	Human CDNA clone (
C	24	23	36.5	2159	22	AAH13744	Human CDNA sequenc
C	25	23	36.5	2263	22	AAH15315	Human CDNA sequenc
C	26	23	36.5	4118	22	AAH161358	Human polynucleoti
C	27	23	36.5	4387	22	AAH159574	Human polynucleoti
C	28	23	36.5	6727	23	ABH23919	Drosophila melanog
C	29	23	36.5	8923	23	ABH23918	Drosophila melanog
C	30	23	36.5	15865	22	AAK90825	Human digestive sy
C	31	23	36.5	910715	20	AAK20248	Borrelia burgdorfe
C	32	22.8	36.2	329	14	AAH061032	Human brain expres
C	33	22.8	36.2	393	23	AAH564499	DNA encoding novel
C	34	22.8	36.2	426	22	AAH134733	Human musculoskele
C	35	22.8	36.2	523	22	AAH05177	Human CDNA clone (
C	36	22.8	36.2	1379	22	AAH81469	Corn promoter clon
C	37	22.8	36.2	2355	24	ABH90356	Human polynucleoti
C	38	22.8	36.2	2499	22	AAH06739	Polynucleotide seq
C	39	22.8	36.2	2553	19	AAV29686	A. thermophilum th
C	40	22.8	36.2	2700	7	AAH60401	Sequence constitut
C	41	22.8	36.2	3767	22	AAH15762	Human CDNA sequenc
C	42	22.8	36.2	8298	14	AAH050168	P190 protein gene.
C	43	22.4	35.6	476	21	AAH43565	Zea mays DNA fragm
C	44	22.4	35.6	564	22	AAH10305	Human CDNA clone (
C	45	22.4	35.6	1275	23	AAH82520	DNA encoding novel

ALIGNMENTS

RESULT	1
AAK59771	
ID	AAK59771 standard; CDNA: 63 BP.
XX	
AC	AAK59771;
XX	
DT	26-JUL-1999 (first entry)
XX	
DE	CDNA encoding a thanatine peptide.
XX	
KW	Thanatine; fungal disease; bacterial disease; Cercospora beticola;
KW	Cladosporium herbarum; Fusarium culmorum; F. graminearum;
KW	Phytophthora cinamomi; selection marker; plant transformation;
KW	herbicide resistance; ss.
XX	
OS	Pseudus sp.
XX	
PN	FR2770853-A1.
XX	
PD	14-MAY-1999.
XX	
PF	07-NOV-1997; 97FR-0014263.
XX	
PR	07-NOV-1997; 97FR-0014263.
XX	
PA	(RHON) RHONE-POULENC AGROCHIMIE.
XX	
PI	Derosse R, Freyssinet G, Hoffmann J;
XX	
DR	WPI, 1999-315645/27.
XX	
DR	P-SDB; AAH15463.
XX	

PT New nucleic acid encoding thanatine useful as a selection marker for
transformation of plants
XX
PS Claim 4; Page 13; 24pp; French.

CC The specification describes a nucleic acid sequence containing the
CC sequence that encodes thanatine. Plants transformed with this nucleic
CC acid sequence are resistant to fungal and bacterial diseases,
CC specifically those caused by *Cercospora beticola*, *Cladosporium*
CC *herbarum*, *Fusarium culmorum*, *F. graminearum* and *Phytophthora*
CC *cinamoni*. The nucleic acid sequence may also be used as a selection
CC marker for transformation of plants with other coding sequences,
CC e.g. those that impart resistance to herbicides. The present sequence
CC encodes the peptide thanatine.

XX
SQ Sequence 63 BP; 19 A; 14 C; 19 G; 11 T; 0 other;

Query Match 100.0%; Score 63; DB 20; Length 63;
Best Local Similarity 100.0%; Pred. No. 2e-13;
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggtccaaagaagccagtcgcaatcattctactgtcaacagagactgttaagtccacagg 60
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1 ggtccaaagaagccagtcgcaatcattctactgtcaacagagagactgttaagtccacagg 60

QY 61 atg 63
|||
Db 61 atg 63

RESULT 2
AAK59772
ID AAK59772 standard; CDNA; 98 BP.

XX
XX
AC AAK59772;

DT 26-JUL-1999 (first entry)

XX
DE cDNA encoding the signal peptide of the PR-1a gene of tobacco.

XX
XX
KM Thanatine; fungal disease; bacterial disease; *Cercospora beticola*;
KM *Cladosporium herbarum*; *Fusarium culmorum*; *F. graminearum*;
KM *Phytophthora cinamoni*; selection marker; plant transformation;
KM herbicide resistance; PR-1a gene; tobacco; ss.

XX
OS *Nicotiana tabacum*.

XX
PN FR2770853-A1.

XX
PD 14-MAY-1999.

XX
PF 07-NOV-1997; 97FR-0014263.

XX
PR 07-NOV-1997; 97FR-0014263.

XX
PA (RHON) RHONE-POULENC AGROCHIMIE.

XX
PI Derose R, Freyssinet G, Hoffmann J;

XX
DR WPI: 1999-315645/27.

XX
DR P-PSDB: AAY15464.

XX
PT New nucleic acid encoding thanatine useful as a selection marker for
transformation of plants

XX
PS Example 1; Page 14; 24pp; French.

XX The specification describes a nucleic acid sequence containing the
CC sequence that encodes thanatine. Plants transformed with this nucleic
CC acid sequence are resistant to fungal and bacterial diseases,
CC specifically those caused by *Cercospora beticola*, *Cladosporium*
CC *herbarum*, *Fusarium culmorum*, *F. graminearum* and *Phytophthora*

CC *cinamoni*. The nucleic acid sequence may also be used as a selection
CC marker for transformation of plants with other coding sequences,
CC e.g. those that impart resistance to herbicides. The present sequence
CC encodes the signal peptide of the PR-1a gene of tobacco, which was
CC used to make fusion products of the invention.

XX
SQ Sequence 98 BP; 25 A; 23 C; 34 G; 16 T; 0 other;

Query Match 100.0%; Score 63; DB 20; Length 98;
Best Local Similarity 100.0%; Pred. No. 2.2e-13;
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggtccaaagaagccagtcgcaatcattctactgtcaacagagactgttaagtccacagg 60
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1 ggtccaaagaagccagtcgcaatcattctactgtcaacagagagactgttaagtccacagg 60

QY 61 atg 63
|||
Db 61 atg 63

RESULT 3
AAK59774
ID AAK59774 standard; CDNA; 197 BP.

XX
XX
AC AAK59774;

DT 26-JUL-1999 (first entry)

XX
DE cDNA encoding a thanatine fusion product of the invention.

XX
XX
KM Thanatine; fungal disease; bacterial disease; *Cercospora beticola*;
KM *Cladosporium herbarum*; *Fusarium culmorum*; *F. graminearum*;
KM *Phytophthora cinamoni*; selection marker; plant transformation;
KM herbicide resistance; PR-1a gene; tobacco; ss.

XX
OS Synthetic.

XX
PN FR2770853-A1.

XX
PD 14-MAY-1999.

XX
PF 07-NOV-1997; 97FR-0014263.

XX
PR 07-NOV-1997; 97FR-0014263.

XX
PA (RHON) RHONE-POULENC AGROCHIMIE.

XX
PI Derose R, Freyssinet G, Hoffmann J;

XX
DR WPI: 1999-315645/27.

XX
DR P-PSDB: AAY15466.

XX
PT New nucleic acid encoding thanatine useful as a selection marker for
transformation of plants

XX
PS Claim 8; Page 14-15; 24pp; French.

XX The specification describes a nucleic acid sequence containing the
CC sequence that encodes thanatine. Plants transformed with this nucleic
CC acid sequence are resistant to fungal and bacterial diseases,
CC specifically those caused by *Cercospora beticola*, *Cladosporium*
CC *herbarum*, *Fusarium culmorum*, *F. graminearum* and *Phytophthora*
CC *cinamoni*. The nucleic acid sequence may also be used as a selection
CC marker for transformation of plants with other coding sequences,
CC e.g. those that impart resistance to herbicides. The present sequence
CC encodes a thanatine fusion product of the invention, comprising the
CC signal peptide of the PR-1a gene of tobacco.

XX
SQ Sequence 197 BP; 32 A; 56 C; 49 G; 60 T; 0 other;

Query Match
Best Local Similarity 100.0%; Score 63; DB 20; Length 197;
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gggtccaaagaacagctgcacatctactgcaacagaggaactggtgaagtcacagag 60
DB 102 gggtccaaagaacagctgcacatctactgcaacagaggaactggtgaagtcacagag 161

QY 61 atg 63
DB 162 atg 164

RESULT 4
AAK59770
ID AAK59770 standard; CDNA; 33 BP.

AC AAK59770;
AT *26-JUL-1999 (first entry)
DE CDNA encoding a thanatine peptide.
KW Thanatine: fungal disease; bacterial disease; Cercospora beticola;
KW Cladosporium herbarum; Fusarium culmorum; F. graminearum;
KW Phytophthora cinamomi; selection marker; plant transformation;
KW herbicide resistance; ss.
OS Psodius sp.
XX FR2770853-A1.
XX PD 14-MAY-1999.
XX PF 07-NOV-1997; 97FR-0014263.
XX PR 07-NOV-1997; 97FR-0014263.
XX PA (RHON) RHONE-POULENC AGROCHIMIE.
XX PI Deroose R, Freyssinet G, Hoffmann J;
XX DR WPI: 1999-315645/27.
XX DR P-PSDB; AAY15462.
PT New nucleic acid encoding thanatine useful as a selection marker for
PT transformation of plants
PS Claim 3; Page 13; 24pp; French.
XX The specification describes a nucleic acid sequence containing the
XX sequence that encodes thanatine. Plants transformed with this nucleic
XX acid sequence are resistant to fungal and bacterial diseases,
XX specifically those caused by Cercospora beticola, Cladosporium
XX herbarum, Fusarium culmorum, F. graminearum and Phytophthora
XX cinamomi. The nucleic acid sequence may also be used as a selection
XX marker for transformation of plants with other coding sequences,
XX e.g. those that impart resistance to herbicides. The present sequence
XX encodes the peptide thanatine.

SQ Sequence 33 BP; 10 A; 7 C; 9 G; 7 T; 0 other;

Query Match
Best Local Similarity 52.4%; Score 33; DB 20; Length 33;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 22 atcatctactgcaacagaggaactggtgaagtc 54
DB 1 atcatctactgcaacagaggaactggtgaagtc 33

RESULT 5

AAH87754/C
ID AAH87754 standard; CDNA; 589 BP.

AC AAH87754;
AT 25-SEP-2001 (first entry)
DE Peppermint plant oil gland expressed cDNA 110.
KW Peppermint; plant oil gland cell; terpenoid essential oil; resin;
KW genetic mapping; antisense suppression; recombinant expression; ss.
OS Mentha x piperita.
XX WO200153319-A1.
XX PD 26-JUL-2001.
XX PF 19-JAN-2001; 2001WO-US02567.
XX PR 20-JAN-2000; 2000US-0177264.
XX PA (CROTEAU R B.
XX PA (LANG/) LANGE B M R.
XX PA (WILD/) WILDUNG M R.
XX PI Croteau RB, Lange BM, Wildung MR;
XX DR WPI: 2001-488706/53.
XX PT New nucleic acid molecules corresponding to mRNA molecules expressed in
XX peppermint oil glands for enhancing expression of plant oil gland cell
XX proteins -
XX PS Claim 1; Page 119-120; 251pp; English.
XX The invention relates to nucleic acid molecules (AAH87645-AAH88116) that
XX correspond to all or part of a mRNA molecule expressed in plant oil
XX gland cells, especially peppermint and plant oil glands that produce
XX terpenoid essential oils and resins. The nucleic acids are useful for
XX genetically mapping a plant genome for genes expressed in plant oil
XX gland cells and to suppress (for example by antisense suppression) or
XX enhance their expression (for example by genetically transforming a
XX plant cell with a replicable expression vector that expresses one or more
XX proteins naturally expressed in plant oil gland cells). The nucleic acids
XX are also useful for recombinant expression of plant oil gland proteins
XX required for terpenoid essential oil and/or resin production in bacterial
XX and/or yeast cells.

SQ Sequence 589 BP; 164 A; 115 C; 135 G; 175 T; 0 other;

Query Match
Best Local Similarity 42.5%; Score 26.8; DB 22; Length 589;
Matches 34; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 9 gaagccagtcacatctactgcaacagaggaactggtgaagtc 54
DB 409 GAGGCCAGTTTATATCATCCACAGACACACACATGCTACAC 364

RESULT 6
AAF12996/C
ID AAF12996 standard; CDNA; 2022 BP.

AC AAF12996;
AT 13-MAR-2001 (first entry)
DE Aspergillus oryzae EST SEQ ID NO:5519.
XX Multiple gene expression; filamentous fungal cell; EST;
XX expressed sequence tag; Fusarium venenatum; Aspergillus niger;

KW Aspergillus oryzae; Trichoderma reesei; Identification; recombination;
KW culture condition; environmental stress; spore morphogenesis;
KW metabolic pathway engineering; catabolic pathway engineering; ss.
XX Aspergillus oryzae.
PN WO200056762-A2.
XX
PD 28-SEP-2000.
XX
PF 22-MAR-2000; 2000WO-US07781.
XX
PR 22-MAR-1999; 99US-0273623.
XX
PA (NOVO) NOVO NORDISK BIOTECH INC.
XX (NOVO) NOVO NORDISK AS.
PI Berka RM, Rey MM, Shuster JR, Kaupinen S, Clausen IG, Olsen PB;
XX WPI; 2000-594572/56.
XX
PT Monitoring differential expression of genes in filamentous fungal cells
PT uses fluorescence-labeled nucleic acids isolated from the cells and a
PT substrate of expressed sequence tags -
XX
PS Claim 88; Page 2292-2293; 3161pp; English.
XX
CC The present invention describes a method for monitoring differential
CC expression of genes in a first filamentous fungal (FF) cell relative to
CC expression of the same genes in one or more second filamentous fungal
CC cells. The method uses fluorescence-labeled nucleic acids isolated from
CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
CC are used in the methods for monitoring differential expression of genes
CC in a first filamentous fungal (FF) cell relative to expression of the
CC same genes in one or more second filamentous fungal cells. Monitoring
CC the global expression of genes from FF cells allows the production
CC potential of the microorganisms to be improved. New genes may be
CC discovered, possible functions of unknown open reading frames can be
CC identified and gene copy number variation and stability can be
CC monitored. The expression of genes can be used to study how FF cells
CC adapt to changes in culture conditions, environmental stress, spore
CC morphogenesis, recombination, metabolic or catabolic pathway
CC engineering. Using ESTs provides several advantages over genomic or
CC random cDNA clones including elimination of redundancy as one spot on an
CC array equals one gene or open reading frame, and organisation of the
CC microarrays based on function of the gene products to facilitate
CC analysis of the results. AAF07478 to AAF11247 represents ESTs from
CC Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus
CC niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and
CC AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
CC all specifically claimed in the present invention.
XX
SQ Sequence 2022 BP; 481 A; 597 C; 483 G; 461 T; 0 other;

Query Match 41.6%; Score 26.2; DB 21; Length 2022;
Best Local Similarity 67.3%; Pred. No. 8.5;
Matches 37; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

OY 9 gaaccagtcgcaatcctactgcaacagagagctgtaattgccaagagatg 63
DB 927 GATGGCAGGCTCAATCTTCACGACACATCGAAGAGGAGTTCAGGGGATG 873

RESULT 7
AAF07491/C
ID AAF07491 standard; cDNA; 1710 BP.
XX
AC AAF07491;
XX
XX 13-MAR-2001 (first entry)
DT
XX Fusarium venenatum EST SEQ ID NO:14.
DE

XX Multiple gene expression; filamentous fungal cell; EST;
KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;
KW Aspergillus oryzae; Trichoderma reesei; Identification; recombination;
KW culture condition; environmental stress; spore morphogenesis;
KW metabolic pathway engineering; catabolic pathway engineering; ss.
XX
XX Fusarium venenatum.
OS
XX
PN WO200056762-A2.
XX
PD 28-SEP-2000.
XX
PF 22-MAR-2000; 2000WO-US07781.
XX
PR 22-MAR-1999; 99US-0273623.
XX
PA (NOVO) NOVO NORDISK BIOTECH INC.
XX (NOVO) NOVO NORDISK AS.
PI Berka RM, Rey MM, Shuster JR, Kaupinen S, Clausen IG, Olsen PB;
XX WPI; 2000-594572/56.
XX
PT Monitoring differential expression of genes in filamentous fungal cells
PT uses fluorescence-labeled nucleic acids isolated from the cells and a
PT substrate of expressed sequence tags -
XX
PS Claim 86; Page 368; 3161pp; English.
XX
CC The present invention describes a method for monitoring differential
CC expression of genes in a first filamentous fungal (FF) cell relative to
CC expression of the same genes in one or more second filamentous fungal
CC cells. The method uses fluorescence-labeled nucleic acids isolated from
CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
CC are used in the methods for monitoring differential expression of genes
CC in a first filamentous fungal (FF) cell relative to expression of the
CC same genes in one or more second filamentous fungal cells. Monitoring
CC the global expression of genes from FF cells allows the production
CC potential of the microorganisms to be improved. New genes may be
CC discovered, possible functions of unknown open reading frames can be
CC identified and gene copy number variation and stability can be
CC monitored. The expression of genes can be used to study how FF cells
CC adapt to changes in culture conditions, environmental stress, spore
CC morphogenesis, recombination, metabolic or catabolic pathway
CC engineering. Using ESTs provides several advantages over genomic or
CC random cDNA clones including elimination of redundancy as one spot on an
CC array equals one gene or open reading frame, and organisation of the
CC microarrays based on function of the gene products to facilitate
CC analysis of the results. AAF07478 to AAF11247 represents ESTs from
CC Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus
CC niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and
CC AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
CC all specifically claimed in the present invention.
XX
SQ Sequence 1710 BP; 395 A; 508 C; 401 G; 399 T; 7 other;

Query Match 39.7%; Score 25; DB 21; Length 1710;
Best Local Similarity 69.4%; Pred. No. 22;
Matches 34; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

OY 14 cagtgccaatcctactgcaacagagagctgtaattgccaagagat 62
DB 624 CAGTCCCAATCTTCATGCGACGATGAGAGGGGAAATTCACGGGAT 576

RESULT 8
AAA91430/C
ID AAA91430 standard; DNA; 1599 BP.
XX
AC AAA91430;
XX

DT	12-JUL-2001	(first entry)
XX	Human CHRM5 coding sequence.	
XX	CHRM5; human; cholinergic receptor muscarinic 5; polymorphic variant;	
KM	genotyping; haplotype; gene therapy; ds.	
XX		
OS	Homo sapiens.	
XX		
FX	Key	
FX	1:1599	
FT	/*tag= a	
FT	/product= "CHRM5"	
FT	replace (26,G)	
FT	/*tag= b	
FT	replace (33,T)	
FT	/*tag= c	
FT	replace (345,T)	
FT	/*tag= d	
FT	replace (573,T)	
FT	/*tag= e	
FT	replace (1234,A)	
FT	/*tag= f	
XX		
PN	*MO200128935-A2.	
XX		
PD	26-APR-2001.	
XX		
PF	19-OCT-2000; 2000MO-US29071.	
XX		
PR	21-OCT-1999; 99US-0160647.	
PA	(GENA-) GENA1SSANCE PHARM INC.	
XX		
PI	Chew A, Chol JY, Nandabalan K, Stephens JC;	
DR	WPI: 2001-300313/31.	
DR	P-PSDB; AAY97746.	
XX		
PT	Isolated polynucleotide encoding the cholinergic receptor, muscarinic 5	
PT	(CHRM5), used to genotype/haplotype the CHRM5 gene, and to identify an	
PT	association between a trait and a polymorphism, comprises novel	
PT	polymorphisms	
XX		
PS	Claim 7; Fig 2; 53pp; English.	
XX		
CC	This sequence encodes the human cholinergic receptor, muscarinic 5	
CC	(CHRM5) protein. The invention relates to a polymorphic variant of this	
CC	CHRM5 sequence. The polymorphic sequence is useful to genotype or	
CC	haplotype the CHRM5 gene, to predict a haplotype pair for the CHRM5 gene,	
CC	and for identifying an association between a trait (such as a clinical	
CC	response to a drug targeting CHRM5). It is also useful in gene therapy in	
CC	patients who lack the CHRM5 isogene or have only one copy of it, and in	
CC	assays to measure the binding affinities of one or more candidate drugs	
CC	targeting CHRM5. The DNA sequence is used in the treatment of disorders	
CC	affected by expression or function of a novel CHRM5 isogene of the	
CC	invention. The protein encoded by the CHRM5 variant is useful to identify	
CC	invention which target the CHRM5 polymorphic variant protein. Antipodles	
CC	against the protein can be used to neutralise the CHRM5 isoform activity	
CC	expressed in an individual, and is useful in detection of CHRM5 in	
CC	immunocytochemical, immunohistochemical and immunofluorescence. A	
CC	composition containing a genotyping oligonucleotide for detecting a	
CC	polymorphism in the CHRM5 gene is used to detect novel CHRM5	
CC	polymorphisms of the invention.	
XX		
SO	Sequence 1599 BP; 407 A; 455 C; 376 G; 357 T; 0 other:	

Query Match 39.0%; Score 24.6; DB 22; Length 1599;

Best Local Similarity 70.2%; Pred. No. 30;

Matches 53; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

17 Tgcacatcactctgacacagagagactgtygaaglyccacagagatg 63

```
Db      523   TCCCAACCAAGTACTGCGCAGCACAGAGAATTGCTGGGCCCCAGAGATG    477  
        | ||||| | | |||||| | |||| | | | | |||||||||  
  
RESULT          9  
ID              AB198000/c  
XX             AB198000 standard; cDNA; 1599 BP.  
AC  
NC             AB198000;  
CD  
DE            18-FEB-2002 (first entry)  
DF  
DM  
DS           Non-endogenous human GPCR cDNA, SEQ ID NO: 520.  
DT  
DW           Human; G protein-coupled receptor; GPCR; non-endogenous; mutant;  
KW           constitutively activated GPCR; agonist; disease; ss.  
XX  
OS           Homo sapiens.  
OC           Synthetic.  
PN           WO200177172-A2.  
PD  
PE            18-OCT-2001.  
PF  
PI            05-APR-2001; 2001WO-US11098.  
PR  
PS            07-APR-2000; 2000US-195747P.  
PT  
PW           (AREN-) ARENA PHARM INC.  
RX  
RY  
SZ  
TA            Lehmann-Bruhnsma K, Liaw CW, Lin I;  
DB            WPI: 2001-648759/74.  
DR  
DI            P-RSDB; ABB56364.  
DX  
DY  
EZ  
FI            Identifying agonists of G protein-coupled receptors (GPCRs) for use in  
FT            disease treatment, comprises contacting candidate compounds with  
FX            versions of GPCRS -  
FY  
FZ  
GA            Example 2; Page 321-322; 394pp; English.  
GC  
GD  
GE           The invention relates to G protein-coupled receptors (GPCRs) for which  
GF           the endogenous ligand has been identified. Non-endogenous  
GH           constitutively activated versions of known GPCRs are used in the  
GI           invention for the direct identification of candidate compounds as  
GJ           receptor agonists, inverse agonists or partial agonists. Such  
GK           agonists are useful as therapeutic agents for diseases or disorders  
GL           associated with GPCRs. The present sequence encodes a non-endogenous  
GM           version of a known human GPCR.  
GN  
GO           Sequence 1599 BP; 410 A; 457 C; 375 G; 357 T; 0 other;  
GP  
GQ  
GR  
GS  
GT  
GU  
GV  
GW  
GX  
GY  
ZZ
```


XX DE DNA encoding novel human diagnostic protein #28401.
XX XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KM Food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
XX 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
PI *Dmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
XX P-PSDB: ABG28410.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1: SEQ ID NO 28401; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (II) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 17745 BP; 4061 A; 5222 C; 4755 G; 3707 T; 0 other;

Query Match 38.7%; Score 24.4; DB 23; Length 17745;
Best Local Similarity 68.0%; Pred. No. 63;
Matches 34; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 5 ccaagaagccagtcacatcatctactgcaacagagagctgtaagtc 54
||| |||| ||| | | | | | | | | | | | | | | | | | | |
Db 17692 CCAAGAGAGAGTGCTGTGAGCTGACGCCACAGCAGACTGTGAGAGC 17643

RESULT 13
ID ABA05165 standard; cDNA; 2604 BP.
XX
AC ABA05165;
XX
DT 04-MAR-2002 (first entry)
XX

DE Human gamma glutamyl transpeptidase coding sequence.
XX
XX Human; gamma glutamyl transpeptidase; glutamate metabolic defect;
KM organic acidemia; metabolic disorder; cancer; inflammation; disease;
KM embryonic developmental disease; growth dysmaturity; immune disease;
KM cytosolic; hemostatic; virucide; immunomodulatory; antiinflammatory;
XX gene therapy; ss.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
FH 2241..2519
FT CDS /tag= a
FT /product= "gamma glutamyl transpeptidase"
XX
XX WO200187944-A1.
XX
PD 22-NOV-2001.
XX
PE 28-APR-2001; 2001WO-CN00660.
XX
PR 29-APR-2000; 2000CN-0115514.
XX
PA (SHAN-) SHANGHAI BIOWINDOW GENE DEV INC.
XX
PI Mao Y, Xie Y;
XX
XX WPI; 2002-055687/07.
XX P-PSDB: AAM47361.
XX
XX Gamma glutamyl transpeptidase and encoding polynucleotide, used in
PT diagnosis and treatment of malignant tumors, hemopathy, human
PT immunodeficiency virus infection, immunological diseases and
PT inflammation -
XX
PS Claim 6: Page 31-32; 39pp; Chinese.
XX
XX The present invention provides the protein and coding sequences of a
CC human gamma glutamyl transpeptidase. The sequences can be used in the
CC treatment of cancer, glutamate metabolic defects, organic acidemia,
CC metabolic disorders, embryonic developmental diseases, growth
CC dysmaturity, inflammation and immune diseases. The present sequence is
CC the coding sequence of the invention.
XX
SQ Sequence 2604 BP; 677 A; 550 C; 776 G; 601 T; 0 other;

Query Match 38.4%; Score 24.2; DB 24; Length 2604;
Best Local Similarity 62.3%; Pred. No. 47;
Matches 38; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 2 gtcccaagaagccagtcacatcatctactgcaacagagagctgtaagtcacagaga 61
||| |||| ||| | | | | | | | | | | | | | | | | | | |
Db 1627 gtccgaagtagcaatgacacacactctttggaagagatgagtgcaagtggaaggga 1666

QY 62 t 62
Db 1687 t 1687

RESULT 14
ID AAQ40706 standard; DNA; 31122 BP.
XX
AC AAQ40706;
XX
DT 17-AUG-1993 (first entry)
XX
DE Bacillus subtilis srfA operon.
XX
XX Multienzyme complex; surfactin synthetase; MCSS; ORF; surfactant;
KM anti-cholinesterase; fungicidal; antibiotic; anti-coagulant;
KM thrombosis; myocardium infarct; pulmonary embolism; ss.

CC (AB857737-AB872072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 4041 BP; 992 A; 1001 C; 1071 G; 977 T; 0 other;

Query Match 37.5%; Score 23.6; DB 23; Length 4041;
Best Local Similarity 69.6%; Pred. No. 87;
Matches 32; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
QY 15 agtgcacatcatctactgtcaacagagagactgtgtaagtcagag 60
I
Db 3021 atggccatttgcactccaccacgacgatgcaatgtgcagtg 3066

Search completed: May 12, 2002, 20:20:39
Job Time: 4926 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 12, 2002, 20:13:05 ; Search time 1854.92 seconds
(without alignments)
710.743 Million cell updates/sec

Title: US-09-554-024-2
Perfect score: 63
Sequence: 1 ggttcaagaagcagtgcc.....ctgtaagtgcagagatgc 63

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:
1: gb_ba:*
2: gb_bt:*
3: gb_in:*
4: gb_lm:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result Query
No. Score Match Length DB ID Description

1	63	100.0	63	6	A95924	A95924	Sequence 2
2	63	100.0	98	6	A95925	A95925	Sequence 3
3	63	100.0	197	6	A95927	A95927	Sequence 5
4	52	82.5	87	6	A95931	A95931	Sequence 9
5	40.4	64.1	42	6	A95930	A95930	Sequence 8
6	33	52.4	33	6	A95923	A95923	Sequence 1
7	27.6	43.8	220257	2	AC005342	AC005342	Homo sapi
8	27.6	43.8	236177	2	AC093660	AC093660	Homo sapi
9	27	42.9	63435	2	AC103764	AC103764	Homo sapi
10	27	42.9	74671	2	AL449106	AL449106	Human DNA
11	27	42.9	168081	2	AC083955	AC083955	Mus muscu
12	27	42.9	176163	2	AL162255	AL162255	Homo sapi
13	27	42.9	192382	2	AC104697	AC104697	Homo sapi
14	27	42.9	197418	30	AC024403	AC024403	Homo sapi
15	26.8	42.5	1632	10	S75713	S75713	abd-2-aceta
16	26.8	42.5	1659	8	AF275347	AF275347	Cladospor
17	26.8	42.5	2038	10	MUSAHDPF	M74570	Mouse aldel
18	26.6	42.2	116181	9	AC010681	AC010681	Homo sapi
19	26.6	42.2	2436	3	AF148508	AF148508	Placoept
20	26.2	41.6	2486	3	AF175578	AF175578	Placoept
21	26.2	41.6	184351	2	AC079783	AC079783	Homo sapi
22	26.2	41.6	190689	2	AC091185	AC091185	Homo sapi
23	26.2	41.6	194543	2	AC027630	AC027630	Homo sapi
24	26	41.3	286550	1	SMES91785	AL591785	Sinorhizo
25	25.8	41.0	106997	2	AC094530	AC094530	Rattus no
26	25.8	41.0	173553	2	AC094561	AC094561	Rattus no
27	25.6	40.6	96489	2	AC021441	AC021441	Homo sapi
28	25.6	40.6	173071	2	AC094125	AC094125	Rattus no
29	25.4	40.3	41631	2	TBCHRIA_09	Continuati	(10 o
30	25.2	40.0	2029	5	AF035443	AF035443	Xenopus 1
31	25.2	40.0	22838	1	AF081920	AF081920	Pseudom
32	25.2	40.0	34158	1	SKL13	AL51267	Streptom
33	25.2	40.0	84399	2	AC094166	AC094166	Rattus no
34	25.2	40.0	137229	9	AC010877	AC010877	Homo sapi
35	25.2	40.0	139737	9	HS228H13	AL031985	Human cDN
36	25.2	40.0	172079	2	AC099744	AC099744	Papio cyn
37	25.2	40.0	173692	2	AC025994	AC025994	Homo sapi
38	25.2	40.0	175509	2	AC099333	AC099333	Homo sapi
39	25.2	40.0	223725	9	AC005820	AC005820	Homo sapi
40	25.2	40.0	258615	2	AC099306	AC099306	Rattus no
41	25	39.7	110000	2	AC084416	Continuati	(4 of
42	25	39.7	122005	9	AC018756	AC018756	Homo sapi
43	25	39.7	124389	2	AP004019	AP004019	Oryza sat
44	25	39.7	139226	9	AC008650	AC008650	Homo sapi
45	25	39.7	150864	9	AC083871	AC083871	Homo sapi

ALIGNMENTS

RESULT 1	A95924	63 bp	DNA	Linear	PAT 26-JAN-2000
LOCUS	A95924	Sequence 2 from Patent WO9924594.			
DEFINITION	A95924				
ACCESSION	A95924.1	GI:6779828			
VERSION					
KEYWORDS					
SOURCE					
ORGANISM					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
DEPOSE					
LOCATION/Qualifiers					
1..63					
/organism="unidentified"					
/db_xref="taxon:32644"					
<1..>63					
/note="unnamed protein product"					
/codon_start=1					

CDS

BASE COUNT 19 a 14 c 19 g 11 t
ORIGIN /protein_id="CAB70492.1"
/db_xref="GI:6779829"
/translation="GSKKPVPIIYCNRRGKOCRM"

Query Match 100.0%; Score 63; DB 6; Length 63;
Best Local Similarity 100.0%; Pred. No. 1.4e-13;
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggttccaagaagccagtcgcaatcattctactgcaacagagagactgtaagtcgcaag 60
|||||
Db 1 GGTTCGAAGAAGCCAGTCGCAATCATCTACTGCAACAGAGAGACTGTAAGTCGCAAG 60

Oy 61 atg 63
|||
Db 61 ATG 63

RESULT 2
LOCUS A95925 98 bp DNA linear PAT 26-JAN-2000
DEFINITION Sequence 3 from Patent WO924594.
ACCESSION A95925
VERSION A95925.1 GI:6779830
KEYWORDS
SOURCE unidentified.
ORGANISM unclassified.

REFERENCE 1 (bases 1 to 98)
AUTHORS Derose, R. and Hoffmann, J.

TITLE GENE CODING FOR THANATIN, VECTOR CONTAINING SAME AND RESULTING
TRANSFORMED DISEASE-RESISTANT PLANTS
JOURNAL Patent: WO 9924594-A 3 20-MAY-1999;
DEROSE RICHARD (FR); HOFFMANN JULES (FR)

FEATURES
source Location/Qualifiers
1..98
/organism="unclassified"
/db_xref="taxon:32644"
CDS
1..66
/note="unnamed protein product"
/codon_start=1
/protein_id="CAB70493.1"
/db_xref="GI:6779831"
/translation="GSKKPVPIIYCNRRGKOCRM"

BASE COUNT 25 a 23 c 34 g 16 t
ORIGIN

Query Match 100.0%; Score 63; DB 6; Length 98;
Best Local Similarity 100.0%; Pred. No. 1.5e-13;
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 ggttccaagaagccagtcgcaatcattctactgcaacagagagactgtaagtcgcaag 60
|||||
Db 1 GGTTCGAAGAAGCCAGTCGCAATCATCTACTGCAACAGAGAGACTGTAAGTCGCAAG 60

Oy 61 atg 63
|||
Db 61 ATG 63

RESULT 3
LOCUS A95927 197 bp DNA linear PAT 26-JAN-2000
DEFINITION Sequence 5 from Patent WO924594.
ACCESSION A95927
VERSION A95927.1 GI:6779834
KEYWORDS
SOURCE unidentified.
ORGANISM unclassified.

REFERENCE 1 (bases 1 to 197)
AUTHORS Derose, R. and Hoffmann, J.
TITLE GENE CODING FOR THANATIN, VECTOR CONTAINING SAME AND RESULTING
TRANSFORMED DISEASE-RESISTANT PLANTS
JOURNAL Patent: WO 9924594-A 5 20-MAY-1999;
DEROSE RICHARD (FR); HOFFMANN JULES (FR)

FEATURES
source Location/Qualifiers
1..197
/organism="unclassified"
/db_xref="taxon:32644"
CDS
12..167
/note="unnamed protein product"
/codon_start=1
/protein_id="CAB70495.1"
/db_xref="GI:6779835"
/translation="MGPVLFSQLPSFLVSTLLFLVISHSCRAGSKKPVPIIYCNRR
TKRCORM"

BASE COUNT 32 a 56 c 49 g 60 t
ORIGIN

Query Match 100.0%; Score 63; DB 6; Length 197;
Best Local Similarity 100.0%; Pred. No. 1.5e-13;
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 ggttccaagaagccagtcgcaatcattctactgcaacagagagactgtaagtcgcaag 60
|||||
Db 102 GGTTCGAAGAAGCCAGTCGCAATCATCTACTGCAACAGAGAGACTGTAAGTCGCAAG 161

Oy 61 atg 63
|||
Db 162 ATG 164

RESULT 4
LOCUS A95931 87 bp DNA linear PAT 26-JAN-2000
DEFINITION Sequence 9 from Patent WO924594.
ACCESSION A95931
VERSION A95931.1 GI:6779839
KEYWORDS
SOURCE unidentified.
ORGANISM unclassified.

REFERENCE 1 (bases 1 to 87)
AUTHORS Derose, R. and Hoffmann, J.

TITLE GENE CODING FOR THANATIN, VECTOR CONTAINING SAME AND RESULTING
TRANSFORMED DISEASE-RESISTANT PLANTS
JOURNAL Patent: WO 9924594-A 9 20-MAY-1999;
DEROSE RICHARD (FR); HOFFMANN JULES (FR)

FEATURES
source Location/Qualifiers
1..87
/organism="unclassified"
/db_xref="taxon:32644"
BASE COUNT 14 a 31 c 21 g 21 t
ORIGIN

Query Match 82.5%; Score 52; DB 6; Length 87;
Best Local Similarity 100.0%; Pred. No. 2.5e-09;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 12 gccagtcgcaatcattctactgcaacagagagactgtaagtcgcaagatg 63
|||||
Db 87 GCCAGTCGCAATCATCTACTGCAACAGAGAGACTGTAAGTCGCAAGATG 36

RESULT 5
LOCUS A95930 42 bp DNA linear PAT 26-JAN-2000
DEFINITION Sequence 8 from Patent WO924594.
ACCESSION A95930
VERSION A95930.1 GI:6779838


```

KEYWORDS
SOURCE      unidentified.
ORGANISM    unidentified.
REFERENCE   1 (bases 1 to 42)
AUTHORS     Derose,R. and Hoffmann,J.
TITLE       DEPOSE,R. AND THANATIN, VECTOR CONTAINING SAME AND RESULTING
            TRANSFORMED DISEASE-RESISTANT PLANTS
JOURNAL     PATENT: WO 9824594-A 8 20-MAY-1999;
            DEROSE RICHARD (FR); HOFFMANN JULES (FR)
FEATURES
Source
BASE COUNT  13 a      12 c      10 g      7 t
ORIGIN

Query Match      64.1%; Score 40.4; DB 6; Length 42;
Best Local Similarity 97.6%; Pred. No. 6.8e-05;
Matches 41: Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ggtccaaagagccagtcacatcatcactcagcaagag 42
|||||
Db 1 ggtccaaagagccagtcacatcatcactcagcaagag 42

RESULT 6
LOCUS      A95923          33 bp      DNA      linear      PAT 26-JAN-2000
DEFINITION Sequence 1 from Patent WO9924594.
ACCESSION  A95923
VERSION     A95923.1 GI:6779826
KEYWORDS
SOURCE      unidentified.
ORGANISM    unidentified.
REFERENCE   1 (bases 1 to 33)
AUTHORS     Derose,R. and Hoffmann,J.
TITLE       GENE CODING FOR THANATIN, VECTOR CONTAINING SAME AND RESULTING
            TRANSFORMED DISEASE-RESISTANT PLANTS
JOURNAL     PATENT: WO 9924594-A 1 20-MAY-1999;
            DEROSE RICHARD (FR); HOFFMANN JULES (FR)
FEATURES
Source
1..33
/organism="unidentified"
/db_xref="taxon:32644"
<1..>33
/note="unnamed protein product"
/codon_start=1
/protein_id="CAB70491.1"
/db_xref="GI:6779827"
/translation="IYCNRRTGKC"
BASE COUNT  10 a      7 c      9 g      7 t
ORIGIN

Query Match      52.4%; Score 33; DB 6; Length 33;
Best Local Similarity 100.0%; Pred. No. 0.047;
Matches 33: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 22 atcattactgcaagagagactgtatgac 54
|||||
Db 1 ATCATCTACTGCACAGAGAGACTGTAGTGC 33

RESULT 7
AC005342/c 220257 bp      DNA      linear      PRI 04-DEC-1998
LOCUS      AC005342
DEFINITION Homo sapiens 12p13.3 PAC RPIc15-1096D14 (Roswell Park Cancer
            Institute Human PAC Library) complete sequence.
ACCESSION  AC005342
VERSION     AC005342.1 GI:3810573

```

```

KEYWORDS
SOURCE      hg.
ORGANISM    human.
REFERENCE   1 (bases 1 to 220257)
AUTHORS     Muzny,D., Aronson,A.D., Bouck,J., Bunac,C., Chang,J., Chen,Z.,
            Ding,Y., Dugan,S., Durbin,J., Forcum,J., Garcia,C., Gorelli,J.H.,
            Gorelli,L.L., Hernandez,J., Jackson,L., Kondogewski,N., Lau,S.,
            Leal,B., Lee,E., Lichtarge,O., Liu,M., Logan,O., Lu,J.,
            Marandel,I., Martinez,C., Merscher,S., Montgomery,K., Oswal,G.,
            Pampell,L.R., Parish,B.J., Perez,L., Rashid,N.D., Renault,B.,
            Rives,C., Scherer,S.E., Shen,H., Simon,M., Vo,Q., Williamson,A.,
            Worley,K.C., Yu,W., Zhou,X., Kucherlapati,R., Nelson,D. and
            Gibbs,R.A.
TITLE       Direct Submission
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 220257)
AUTHORS     Worley,K.C.
TITLE       Direct Submission
JOURNAL     Submitted (31-JUL-1998) Molecular and Human Genetics, Baylor
            College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3 (bases 1 to 220257)
AUTHORS     Worley,K.C.
TITLE       Direct Submission
JOURNAL     Submitted (30-OCT-1998) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
4 (bases 1 to 220257)
Worley,K.C.
Direct Submission
Submitted (04-DEC-1998) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Oct 30, 1998 this sequence version replaced g1:3723954.
On Oct 30, 1998 this sequence version replaced g1:3723954.
INFORMATION: http://gc.bcm.tmc.edu:8088/home.html or email
gc-help@bcm.tmc.edu

```

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:
STSs are identified using ePCR (genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished) for Human and Mouse sequences.
Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-14) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as low coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAT-REPORT-----

----- Summary Statistics -----
Contig Length: 220257
Phrap values in estimate: 58222
Average error rate (RCM-Phrap estimate): 9.8888e-07
Fraction of Phrap values less than 40 : 0.00173474
Number of Phrap values changing edits: 10
Number of N's in consensus : 0

----- Consensus Changing Edits -----
Position Original+Context Edited+Context
33512 attcttattat(t)atttatttta atttcttatt(t)atttatttta
54816 aacaaaga(n)aataccaga aacaaaga(a)aataccaga
56795 cgggtgttaac(n)caaggctgga cgggtgttaac(c)caaggctgga
120971 aaaaaaaa(c)caacaactc aaaaaaaa(a)caacaactc
152487 lttcaaaaga(n)actltagaag lttcaaaaga(a)actltagaag
176875 ggggtctgcaag(n)ccggcgcgag ggggtctgcaag(g)ccggcgcgag
176918 cgcgaaggag(n)gttcgcgcgc cgcgaaggag(t)gttcgcgcgc
177533 ccgcgcctc(n)ltccgcctc ccgcgcctc(c)ltccgcctc
178001 ccgcgtccc(n)gcggcccg ccgcgtccc(c)gcggcccg
195474 cctcctccc(n)lccctccc cctcctccc(c)lccctccc

----- Low Quality Bases -----
Quality Position Surrounding Sequence

----- Distribution of Quality < 40 Bases -----

#	5	10	15	20	25	30	35	40
1001								*
901								*
801								*
701								*
601								*
501								*
401								*
301								*
201								*
101								*
01				*	*	*	*	*

	5	10	15	20	25	30	35	40

	Phrap Value Range							

----- FEATURES -----
Version: 1.0 txio. Location/Qualifiers

source 1..220257 /organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="12p13.3"
/clone="RPC15-1096D14"
misc.feature 1..78 /note="Bases 1 to 16417 overlaps the adjacent clone
RPC11-21K20 only 78 bases of overlap included in
AC005343."
/function="Overlap with adjacent clone AC005343."
misc_difference 2593 /note="Insertion of A."
/clone="Insertion with respect to AC005343."
repeat_region complement(3901..4051)
/rpt_family="MER58"
misc_difference 6497 /note="A in AC005343."
/clone="A in AC005343."
repeat_region 7931..7979 /note="Substitution with respect to AC005343."
/rpt_family="L2"
repeat_region 8136..8227 /note="Substitution with respect to AC005343."
/rpt_family="L2"

misc_difference 8487 /note="C in AC005343."
/clone="Substitution with respect to AC005343."
misc_difference 10553 /note="C in AC005343."
/clone="Substitution with respect to AC005343."
misc_difference 10772 /note="C in AC005343."
/clone="Substitution with respect to AC005343."
misc_difference 12486 /note="G in AC005343."
/clone="Substitution with respect to AC005343."
repeat_region 13057..13629 /note="Substitution with respect to AC005343."
/rpt_family="L1M4"
misc_difference 13194 /note="G in AC005343."
/clone="Substitution with respect to AC005343."
misc_difference 13419 /note="C in AC005343."
/clone="Substitution with respect to AC005343."
repeat_region complement(13813..13922) /note="Substitution with respect to AC005343."
/rpt_family="FLAM_A"
repeat_region 14162..17576 /note="Substitution with respect to AC005343."
/rpt_family="L1M4"
misc_difference 14448 /note="C in AC005343."
/clone="Substitution with respect to AC005343."
misc_difference 15248 /note="G in AC005343."
/clone="Substitution with respect to AC005343."
misc_difference 16418 /note="G in AC005343."
/clone="Substitution with respect to AC005343."
misc_difference 16421 /note="G in AC005343."
/clone="Substitution with respect to AC005343."
misc_difference 16431 /note="G in AC005343."
/clone="Substitution with respect to AC005343."
misc_difference 16431 /note="G in AC005343."
/clone="Substitution with respect to AC005343."
misc_difference 16464 /note="G in AC005343."
/clone="Substitution with respect to AC005343."
misc_difference 17577 /note="G in AC005343."
/clone="Substitution with respect to AC005343."
repeat_region 17855..19931 /note="Substitution with respect to AC005343."
/rpt_family="L1M1"
repeat_region complement(19962..20254) /note="Substitution with respect to AC005343."
/rpt_family="AluIdo"
repeat_region 20261..21347 /note="Substitution with respect to AC005343."
Query Match 43.8%; Score 27.6; DB 9; Length 220257;
Best Local Similarity 67.2%; Pred. No. 12;
Matches 39; Conservative 0; Mismatches 19; Indels 0; Gaps 0;
Qy 2 gttccaagaagccagtgccatcatcactgcaacagagagactgttaagccagag 59
Db 98878 GCTCTTAGTTGCGAGGCTTAACAAGTAGTCAATGCAAGAGATGTATATGCTAGAG 98821
RESULT 8
AC093660 236177 bp DNA linear HTG 24-JAN-2002
LOCUS AC093660
DEFINITION Homo sapiens chromosome 4 clone RP11-470D11, WORKING DRAFT
SEQUENCE 23 unordered pieces.
ACCESSION AC093660
VERSION AC093660.2 GI:18308900
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 236177)
Waterson,R.H.

TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 236177)
AUTHORS Waterston, R.H.
TITLE Direct Submission
Submitted (07-SEP-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Jan 24, 2002 this sequence version replaced gi:15487495.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@wustl.wustl.edu
----- Project Information -----
Center project name: H.NH0470D11

----- Summary Statistics -----
Sequencing vector: M13; 0%
Sequencing vector: plasmid; 100%
Chemistry: Dye-terminator Big Dye; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 228670 bases at least Q40
Consensus quality: 23189 bases at least Q30
Consensus quality: 232856 bases at least Q20
Insert size: 171000; agarose-fp
Insert size: 234376; sum-of-contigs
Quality coverage: 11.97 in Q20 bases; agarose-fp
Quality coverage: 8.83 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 23 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1
1169: contig of 1169 bp in length
1170 1269: gap of unknown length
1270 1395: contig of 1126 bp in length
2396 2495: gap of unknown length
2496 3610: contig of 1115 bp in length
3611 3710: gap of unknown length
3711 4819: contig of 1109 bp in length
4820 4919: gap of unknown length
4920 6071: contig of 1152 bp in length
6072 6171: gap of unknown length
6172 7831: contig of 1660 bp in length
7832 7931: gap of unknown length
7932 9100: contig of 1169 bp in length
9101 9200: gap of unknown length
9201 10495: contig of 1295 bp in length
10496 10595: gap of unknown length
10596 11673: contig of 1078 bp in length
11674 11773: gap of unknown length
11774 13026: contig of 1253 bp in length
13027 13126: gap of unknown length
13127 14937: contig of 1811 bp in length
14938 15037: gap of unknown length
15038 16737: contig of 1700 bp in length
16738 16837: gap of unknown length
16838 18422: contig of 1585 bp in length
18423 18522: gap of unknown length
18523 19834: contig of 1312 bp in length
19835 19934: gap of unknown length
19935 22586: contig of 2652 bp in length
22587 22686: gap of unknown length
22687 25068: contig of 2382 bp in length
25069 25168: gap of unknown length
25169 27681: contig of 2513 bp in length

FEATURES
source * 27682 27781: gap of unknown length
* 27782 29591: contig of 1810 bp in length
* 29592 29691: gap of unknown length
* 29692 38111: contig of 8420 bp in length
* 38112 38211: gap of unknown length
* 38212 47759: contig of 9548 bp in length
* 47760 47859: gap of unknown length
* 47860 61458: contig of 13599 bp in length
* 61459 61558: gap of unknown length
* 61559 235391: contig of 173833 bp in length
* 235392 235491: gap of unknown length
* 235492 236177: contig of 686 bp in length.
Location/Qualifiers
1. .236177
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"
/clone="RP11-470D11"
1. .1169
/note="assembly_name:Contig44"
1270. .2395
/note="assembly_name:Contig50"
2496. .3610
/note="assembly_name:Contig56"
3711. .4819
/note="assembly_name:Contig62"
4920. .6071
/note="assembly_name:Contig63"
6172. .7831
/note="assembly_name:Contig65"
7932. .9100
/note="assembly_name:Contig68"
9201. .10495
/note="assembly_name:Contig70"
10596. .11673
/note="assembly_name:Contig74"
11774. .13026
/note="assembly_name:Contig75"
13127. .14937
/note="assembly_name:Contig76"
15038. .16737
/note="assembly_name:Contig77"
16838. .18422
/note="assembly_name:Contig78"
18523. .19834
/note="assembly_name:Contig79"
19935. .22586
/note="assembly_name:Contig80"
22687. .25068
/note="assembly_name:Contig81"
25169. .27681
/note="assembly_name:Contig82"
27782. .29591
/note="assembly_name:Contig84"
29692. .38111
/note="assembly_name:Contig85"
38212. .47759
/note="assembly_name:Contig86"
47860. .61458
/note="assembly_name:Contig87"
61559. .235391
/note="assembly_name:Contig88
clone_end:T7
vector_side:left"
235492. .236177
/note="assembly_name:Contig4"
BASE COUNT 64649 a 51668 c 51720 g 65858 t 2282 others
ORIGIN
Query Match 43.8%; Score 27.6; DB 2; Length 236177;
Best Local Similarity 67.2%; Pred. No. 13;
Matches 39; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

0Y 2 gttccaaagaagccagtgccatcatctactgcacagaggaactgtaagtcgcagag 59
DB 6441 GCTCTTAGTTCGACGCTCCATACCAAGTAGTCATCAAGAGAAATGATATATGTGTAGAG 6498
RESULT 9
AC103764/C
LOCUS
DEFINITION
AC103764 63435 bp DNA linear HTG 29-NOV-2001
SEQUENCE SAMPLING:
AC103764
AC103764.1 GI:17149649
HTG; HTGS_PHASE0.
SOURCE
ORGANISM
Human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
1 (bases 1 to 63435)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, A., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Boguski, L., Bouckgeater, B.,
Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,
Choquet, Y., Collangelo, M., Collins, S., Collymore, A., Cook, A.,
Cooke, P., Deatellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S.,
Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Garlinda, S.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Heald, A., Horton, L., Hulme, W., Illey, I., Johnson, R.,
Jones, C., Kamat, A., Karas, A., Kells, C., Larocque, K.,
Lamacan, C., Macdonald, P., Major, J., Margulis, N., Matthews, C.,
McCarthy, M., McEwan, P., McKernan, K., McPherson, R., Meldrum, J.,
Meneus, L., Milova, T., Mlenka, V., Murphy, T., Naylor, J., Nguyen, C.,
Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,
Oliver, J., Peterson, K., Phukhang, P., Pierre, N., Pollara, V.,
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R.,
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Testa, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Triggillo, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,
Zaloun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (29-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
TITLE
JOURNAL
COMMENT
Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: MIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
Project Information
Center project name: 1057_B_8
Center clone name: 1057_B_8
* NOTE: This record contains 83 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
* 1
* 656 755: gap of 100 bp
* 656 755: gap of 100 bp in length
* 756 1426: contig of 671 bp in length
* 1427 1526: gap of 100 bp
* 1527 2169: contig of 643 bp in length
* 2170 2269: gap of 100 bp
* 2270 2943: contig of 674 bp in length
* 2944 3043: gap of 100 bp
* 3044 3713: contig of 670 bp in length
* 3714 3813: gap of 100 bp
* 3814 4454: contig of 641 bp in length
* 4455 4554: gap of 100 bp
* 4555 5219: contig of 665 bp in length
* 5220 5319: gap of 100 bp
* 5320 5979: contig of 660 bp in length
* 5980 6079: gap of 100 bp
* 6080 6731: contig of 652 bp in length
* 6732 6831: gap of 100 bp
* 6832 7511: contig of 680 bp in length
* 7512 7611: gap of 100 bp
* 7612 8289: contig of 678 bp in length
* 8290 8389: gap of 100 bp
* 8390 9056: contig of 667 bp in length
* 9057 9156: gap of 100 bp
* 9157 9820: contig of 664 bp in length
* 9821 9920: gap of 100 bp
* 9921 10585: contig of 665 bp in length
* 10586 10685: gap of 100 bp
* 10686 11354: contig of 669 bp in length
* 11355 11454: gap of 100 bp
* 11455 12115: contig of 661 bp in length
* 12116 12215: gap of 100 bp
* 12216 12895: contig of 680 bp in length
* 12896 12995: gap of 100 bp
* 12996 13648: contig of 653 bp in length
* 13649 13748: gap of 100 bp
* 13749 14419: contig of 671 bp in length
* 14420 14519: gap of 100 bp
* 14520 15201: contig of 682 bp in length
* 15202 15301: gap of 100 bp
* 15302 15927: contig of 626 bp in length
* 15928 16027: gap of 100 bp
* 16028 16713: contig of 686 bp in length
* 16714 16813: gap of 100 bp
* 16814 17478: contig of 665 bp in length
* 17479 17578: gap of 100 bp
* 17579 18233: contig of 655 bp in length
* 18234 18333: gap of 100 bp
* 18334 19006: contig of 673 bp in length
* 19007 19106: gap of 100 bp
* 19107 19776: contig of 670 bp in length
* 19777 19876: gap of 100 bp
* 19877 20551: contig of 675 bp in length
* 20552 20651: gap of 100 bp
* 20652 21327: contig of 676 bp in length
* 21328 21427: gap of 100 bp
* 21428 22090: contig of 663 bp in length
* 22091 22190: gap of 100 bp
* 22191 22881: contig of 691 bp in length
* 22882 22981: gap of 100 bp
* 22982 23627: contig of 646 bp in length
* 23628 23727: gap of 100 bp
* 23728 24397: contig of 670 bp in length
* 24398 24497: gap of 100 bp
* 24498 25160: contig of 663 bp in length
* 25161 25260: gap of 100 bp
* 25261 25941: contig of 681 bp in length
* 25942 26041: gap of 100 bp
* 26042 26692: contig of 651 bp in length
* 26693 26792: gap of 100 bp
* 26793 27459: contig of 667 bp in length
* 27460 27559: gap of 100 bp
* 27560 28227: contig of 666 bp in length
* 28228 28327: gap of 100 bp
* 28328 28976: contig of 649 bp in length

/note="24 copies 2 mer tt 77% conserved"
repeat_region 2336. .2635
/note="AluX repeat: matches 1. .310 of consensus"
repeat_region 2642. .2924
/note="L2 repeat: matches 2425. .2710 of consensus"
2925. .3332
/note="MIR repeat: matches 1. .426 of consensus"
repeat_region 3333. .3373
/note="L2 repeat: matches 2709. .2749 of consensus"
4952. .5228
/note="AluX repeat: matches 6. .281 of consensus"
repeat_region 5230. .5309
/note="40 copies 2 mer ac 87% conserved"
repeat_region 5489. .5533
/note="L2 repeat: matches 2608. .2656 of consensus"
6092. .6595
/note="L2 repeat: matches 2242. .2710 of consensus"
6596. .6887
/note="AluJ repeat: matches 1. .292 of consensus"
repeat_region 6888. .7350
/note="L2 repeat: matches 1639. .2242 of consensus"
8472. .8710
/note="L2 repeat: matches 7688. .7955 of consensus"
8743. .8826
/note="L2 repeat: matches 5696. .5780 of consensus"
8827. .9138
/note="AluX repeat: matches 1. .312 of consensus"
9139. .9537
/note="L2 repeat: matches 5780. .6169 of consensus"
9537. .9675
/note="L2 repeat: matches 7553. .7665 of consensus"
9682. .9808
/note="MER1 repeat: matches 1. .113 of consensus"
9810. .10079
/note="L2 repeat: matches 7276. .7556 of consensus"
10618. .10975
/note="L2 repeat: matches 1. .354 of consensus"
1111. .11295
/note="L2 repeat: matches 2562. .2750 of consensus"
11308. .11389
/note="L2 repeat: matches 6092. .6173 of consensus"
11753. .12058
/note="AluX repeat: matches 1. .310 of consensus"
12691. .12969
/note="AluX repeat: matches 24. .301 of consensus"
13263. .13312
/note="25 copies 2 mer tt 80% conserved"
14143. .14223
/note="MIR repeat: matches 21. .101 of consensus"
14376. .14437
/note="31 copies 2 mer at 87% conserved"
14439. .14725
/note="AluX repeat: matches 1. .290 of consensus"
14759. .15046
/note="AluS repeat: matches 1. .287 of consensus"
15060. .15489
/note="L2 repeat: matches 203. .630 of consensus"
15512. .15579
/note="L2 repeat: matches 857. .783 of consensus"
15585. .15757
/note="L2 repeat: matches 968. .1138 of consensus"
15765. .15978
/note="L2 repeat: matches 558. .594 of consensus"
15993. .16688
/note="L2 repeat: matches 1424. .2085 of consensus"
16689. .16998
/note="AluS repeat: matches 1. .309 of consensus"
16999. .17455
/note="L2 repeat: matches 2085. .2520 of consensus"
17456. .17768
/note="AluS repeat: matches 1. .312 of consensus"
17769. .18106
/note="L2 repeat: matches 2520. .2841 of consensus"

repeat_region 18107. .18420
/note="AluX repeat: matches 1. .312 of consensus"
repeat_region 18421. .18905
/note="L2 repeat: matches 2841. .3108 of consensus"
18906. .19183
/note="AluX repeat: matches 1. .301 of consensus"
19184. .19423
/note="L2 repeat: matches 3108. .3336 of consensus"
19429. .19639
/note="L2 repeat: matches 3543. .3749 of consensus"
19640. .19757
/note="AluX repeat: matches 1. .112 of consensus"
19758. .20061
/note="AluX repeat: matches 1. .304 of consensus"
20062. .20219
/note="AluX repeat: matches 112. .291 of consensus"
20220. .21445
/note="L2 repeat: matches 3749. .5003 of consensus"
21501. .21625
/note="AluS repeat: matches 170. .296 of consensus"
21653. .21701
/note="L2 repeat: matches 174. .222 of consensus"
21702. .22015
/note="AluX repeat: matches 1. .303 of consensus"
22016. .22489
/note="L2 repeat: matches 491. .962 of consensus"
22490. .22930
/note="MIR repeat: matches 1. .426 of consensus"
22931. .24127
/note="L2 repeat: matches 962. .2175 of consensus"
24128. .24433
/note="AluS repeat: matches 1. .313 of consensus"
24434. .24992
/note="L2 repeat: matches 2175. .2731 of consensus"
24993. .25457
/note="L2 repeat: matches 5682. .6060 of consensus"
25458. .25755
/note="AluX repeat: matches 1. .300 of consensus"
25756. .25991
/note="L2 repeat: matches 6060. .6307 of consensus"
26087. .26393
/note="L2 repeat: matches 2417. .2746 of consensus"
26479. .26679
/note="MIR repeat: matches 24. .255 of consensus"
27620. .27688
/note="L2 repeat: matches 2681. .2749 of consensus"
27641. .27707
/note="MIR repeat: matches 182. .248 of consensus"
27700. .27822
/note="L2 repeat: matches 2624. .2750 of consensus"
27823. .28152
/note="AluX repeat: matches 1. .309 of consensus"

Query Match 42.9%; Score 27; DB 9; Length 74671;
Best Local Similarity 70.6%; Pred.No.19;
Matches 36; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 2 gttccagaagccagtgccatcatctacgcaacaggaagcgtgaagt 52
Db 50221 gttccagaagcaggttccatgtcttttgacacagatcattgtgact 50271

RESULT 11
AC083955 168081 bp DNA linear HTG 22-NOV-2000
LOCUS AC083955/c
DEFINITION Mus musculus clone R23-84P17, WORKING DRAFT SEQUENCE, 36 unordered
pieces:
AC083955
AC083955.2 GI:11276220
VERSION AC083955.2
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.

SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE 1 (bases 1 to 168081)
AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.
JOURNAL Mus musculus chromosome, clone RP23-84P17
REFERENCE Unpublished
AUTHORS 2 (bases 1 to 168081)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Barn, N., Bastien, V., Bede, F., Boguslavsky, L.,
Boutgalter, B., Brown, A., Burkett, G., Campopiano, A., Casle, A.,
Choepe, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P.,
D'Amico, K., Dewar, K., Diaz, J., Dodge, S., Ferreira, P.,
Fitzhugh, W., Gage, D., Galagan, J., Gao, Y., Ginde, S., Goyette, M.,
Graham, L., Grand-Pierre, N., Hagos, B., Heatford, A., Horton, L.,
Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Latocque, K.,
Lamarez, R., Landers, T., Lehotzky, J., Levine, R., Lieu, C., Liu, G.,
Macdonald, P., Margulis, N., McCarthy, M., McEwan, P., McKernan, K.,
McPheeters, R., Meldrim, J., Menus, L., Mihova, T., Mlenga, V.,
Morrow, J., Murphy, T., Naylor, J., Norman, C.H., O'Connor, T.,
O'Donnell, P., O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K.,
Pierre, N., Pisan, C., Pollara, V., Raymond, C., Rieback, M., Riley, R.,
Rogor, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P.,
Sougnuez, C., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Sudramanian, A., Talamas, J., Testaye, S., Theodore, J.,
Tirrell, A., Travers, M., Trifillo, J., Vassiliev, H., Viel, R., Vo, A.,
Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zaitoun, J.,
Zimmer, A. and Zody, M.
Direct Submission
Submitted (08-Oct-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Nov 22, 2000 this sequence version replaced g1:10717222.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www.seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: l11090
Center clone name: 84_P_17
----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 147448 bases at least Q40
Consensus quality: 158296 bases at least Q30
Consensus quality: 162516 bases at least Q20
Insert size: 198000; agarose-tp
Insert size: 164581; sum-of-ctnigs
Quality coverage: 2.8 in Q20 bases; agarose-tp
Quality coverage: 3.4 in Q20 bases; sum-of-ctnigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 36 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 8071: contig of 8071 bp in length
* 8072 8171: gap of 100 bp
* 8172 8798: contig of 627 bp in length
* 8799 8898: gap of 100 bp
* 8899 9551: contig of 653 bp in length
* 9552 9651: gap of 100 bp
* 9652 10408: contig of 757 bp in length
* 10409 10508: gap of 100 bp
* 10509 11835: contig of 1327 bp in length

11836 11935: gap of 100 bp
11936 13013: contig of 1078 bp in length
13014 13113: gap of 100 bp
13114 14364: contig of 1271 bp in length
14385 14484: gap of 100 bp
14485 15189: contig of 705 bp in length
15190 15289: gap of 100 bp
15290 16626: contig of 1337 bp in length
16627 16726: gap of 100 bp
16727 17872: contig of 1146 bp in length
17873 17972: gap of 100 bp
17973 19113: contig of 1141 bp in length
19114 19213: gap of 100 bp
19214 20332: contig of 1119 bp in length
20333 20432: gap of 100 bp
20433 22596: contig of 2164 bp in length
22597 22696: gap of 100 bp
22697 25492: contig of 2796 bp in length
25493 25592: gap of 100 bp
25593 28554: contig of 2962 bp in length
28555 28654: gap of 100 bp
28655 31898: contig of 3244 bp in length
31899 31998: gap of 100 bp
31999 35948: contig of 3850 bp in length
35949 35948: gap of 100 bp
35950 39030: contig of 3082 bp in length
39031 39130: gap of 100 bp
39131 42611: contig of 3461 bp in length
42612 42711: gap of 100 bp
42712 45681: contig of 2970 bp in length
45682 45781: gap of 100 bp
45782 48839: contig of 3058 bp in length
48840 48939: gap of 100 bp
48940 53014: contig of 4075 bp in length
53015 53114: gap of 100 bp
53115 56891: contig of 3777 bp in length
56892 56991: gap of 100 bp
56992 62353: contig of 5362 bp in length
62354 62453: gap of 100 bp
62454 66132: contig of 3679 bp in length
66133 66232: gap of 100 bp
66233 70771: contig of 4539 bp in length
70772 70871: gap of 100 bp
70872 76244: contig of 5373 bp in length
76245 76344: gap of 100 bp
76345 80719: contig of 4375 bp in length
80720 80819: gap of 100 bp
80820 85869: contig of 5050 bp in length
85870 85969: gap of 100 bp
85970 92396: contig of 6427 bp in length
92397 92496: gap of 100 bp
92497 99048: contig of 6552 bp in length
99049 99148: gap of 100 bp
99149 109053: contig of 9905 bp in length
109054 109153: gap of 100 bp
109154 120325: contig of 11172 bp in length
120326 120425: gap of 100 bp
120426 132250: contig of 11825 bp in length
132251 132350: gap of 100 bp
132351 145705: contig of 13355 bp in length
145706 145805: gap of 100 bp
145806 168081: contig of 22276 bp in length.
Location/Qualifiers
1.168081
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="RP23-84P17"
/clone_lib="RP23-84P17"
1.8071
/note="assembly-fragment
clone_end:5p6
vector_side:left"
8172..8798
misc_feature
misc_feature
misc_feature

```
misc_feature      /note="assembly-fragment"
8899. .9551
/misc_feature      /note="assembly-fragment"
9652. .10408
/misc_feature      /note="assembly-fragment"
10509. .11835
/misc_feature      /note="assembly-fragment"
11936. .13013
/misc_feature      /note="assembly-fragment"
13114. .14384
/misc_feature      /note="assembly-fragment"
14485. .15189
/misc_feature      /note="assembly-fragment"
15290. .16626
/misc_feature      /note="assembly-fragment"
16727. .17872
/misc_feature      /note="assembly-fragment"
17973. .19113
/misc_feature      /note="assembly-fragment"
19214. .20332
/misc_feature      /note="assembly-fragment"
20433. .22596
/misc_feature      /note="assembly-fragment"
22697. .25492
/misc_feature      /note="assembly-fragment"
25593. .28554
/misc_feature      /note="assembly-fragment"
28655. .31898
/misc_feature      /note="assembly-fragment"
31999. .35848
/misc_feature      /note="assembly-fragment"
35949. .39030
/misc_feature      /note="assembly-fragment"
39131. .42611
/misc_feature      /note="assembly-fragment"
42712. .45681
/misc_feature      /note="assembly-fragment"
45782. .48839
/misc_feature      /note="assembly-fragment"
48940. .53014
/misc_feature      /note="assembly-fragment"
53115. .56891
/misc_feature      /note="assembly-fragment"
56992. .62353
/misc_feature      /note="assembly-fragment"
62454. .66132
/misc_feature      /note="assembly-fragment"
66233. .70771

Query Match      42.9%: Score 27; DB 2; Length 168081;
Best Local Similarity 70.6%; Pred. No. 21;
Matches 36; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

OY      3      ttccaagaagccagtcgaatcatctactgcacaagagagactgtaagt 53
Db 164047 ttgatagAACGCTGGCGCCATTCATCTGTAAGGAGGAAAGGAGGAGTG 163997

RESULT 12
AL162255      176163 bp      DNA      linear      HTG 12-FEB-2002
LOCUS      Homo sapiens chromosome 1 clone RP11-18E13 map q24.3-25.3, ***
DEFINITION      SEQUENCING IN PROGRESS ***, in ordered pieces.
ACCESSION      AL162255
VERSION      AL162255.16 GI:18655863
KEYWORDS      HTG; HTGS_PHASE2; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE      1 (bases)
AUTHORS      Harrison, E.
TITLE      Direct Submission
```

```
JOURNAL      Submitted (11-FEB-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
On Feb 13, 2002 this sequence version replaced g1:18250452.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
----- Project Information
Center project name: BA18E13
----- Summary Statistics
Assembly program: XGAP4; Version 4.5
Sequencing vector: plasmid; 108752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 175707 bases at least Q40
Consensus quality: 175950 bases at least Q30
Consensus quality: 176098 bases at least Q20
Insert size: 176163; sum-of-contigs
Insert coverage: 172792; agarose-fp
Quality coverage: 7.21x in Q20 bases; sum-of-contigs quality
coverage: 7.90x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
Location/Qualifiers
1. 176163
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/map="q24.3-25.3"
/clone="RP11-18E13"
/clone_1fb="RPC1-11.1"
1. 176163
/note="assembly-fragment:02749
clone.end..f7
vector_side:left"

BASE COUNT      51776 a 32454 c 34988 g 56945 t
ORIGIN

Query Match      42.9%: Score 27; DB 2; Length 176163;
Best Local Similarity 70.6%; Pred. No. 21;
Matches 36; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

OY      2      gtccaagaagccagtcgaatcatctactgcacaagagagactgtaagt 52
Db 29957 GTTCAAGTACACACTTCATCTGTCGACAGGATCTTGTAAAGT 30007

RESULT 13
AC104697      192382 bp      DNA      linear      HTG 19-DEC-2001
LOCUS      Homo sapiens chromosome 2 clone RP11-454B3, WORKING DRAFT SEQUENCE,
2 unordered pieces.
ACCESSION      AC104697 AC024403
VERSION      AC104697.1 GI:17933873
KEYWORDS      HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE      1 (bases 1 to 192382)
AUTHORS      Waterston, R.H.
TITLE      The sequence of Homo sapiens clone
JOURNAL      Unpublished
REFERENCE      2 (bases 1 to 192382)
AUTHORS      Waterston, R.H.
TITLE      Direct Submission
JOURNAL      Submitted (19-DEC-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
```


COMMENT MO 63108, USA
On Dec 19, 2001 this sequence version replaced gi:7229914.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc/index.shtml>
Contact: submissions@genome.wustl.edu
----- Project Information -----
Center project name: H_NH0454B03
Drafting center: WIBR

----- Summary Statistics -----
Sequencing vector: M13; 43%
Sequencing vector: plasmid; 57%
Chemistry: Dye-terminator Big Dye; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 190293 bases at least Q40
Consensus quality: 190786 bases at least Q30
Consensus quality: 191359 bases at least Q20
Insert size: 171000; agarose-fp
Insert size: 192282; sum-of-contigs
Quality coverage: 10.68 in Q20 bases; sum-of-contigs
Quality coverage: 9.51 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 2089: contig of 2089 bp in length
* 2090 2189: gap of unknown length
* 2190 192382: contig of 190193 bp in length.
Location/Qualifiers
1. 192382

FEATURES
source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/clone="RP11-454B3"

misc_feature
1..2089
/note="assembly_name:Contig16"
misc_feature
2190..192382
/note="assembly_name:Contig17
clone_end:SP6
vector_side:right"

BASE COUNT 63187 a 34073 c 34687 g 60334 t 101 others
ORIGIN

Query Match 42.9%; Score 27; DB 2; Length 192382;
Best local Similarity 76.7%; Pred. No. 21;
Matches 33; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

OY 3 ttccaagaagccagtcgcatcattctcaacagagagact 45
|||||
Db 51376 TTCTAGGAAATGCGTACTATCTATCTATCATCAGGACT 51418
|||||

RESULT 14
AC024403/c standard; DNA; HTG; 197418 BP.
XX AC024403;
AC AC024403;
SV AC024403.2
XX
DT 03-MAR-2000 (Rel. 62, Created)
XX 15-MAR-2000 (Rel. 63, last updated, Version 2)
XX

DE Homo sapiens chromosome 2 clone RP11-454B3 map 2, WORKING DRAFT SEQUENCE,
DE 44 unordered pieces.
XX HTG; HTGS_DRAFT; HTGS_PHASE1.
XX
XX Homo sapiens (human)
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominiidae; Homo.
XX
XX [1]
RP 1-197418
RA Birren B., Linton L., Nusbaum C., Lander E.,
RT "Homo sapiens chromosome 2, clone RP11-454B3",
XX unpublished.
XX [2]
RP 1-197418
RA Birren B., Linton L., Nusbaum C., Lander E.,
RA Anderson S., Baldwin J., Barna N., Bede F., Boguslavskiy L., Boukhalter B.,
RA Brown A., Burkett G., Campopiano A., Castle A., Choepel Y., Colangelo M.,
RA Collins S., Collamore A., Cooke P., Dekrellano K., Dewar K., Diaz J.S.,
RA Dodge S., Domino M., Doyle M., Fenesfor J., Ferreira P., Fitzhugh W.,
RA Forrest C., Gage D., Galagan J., Gardyna S., Ginde S., Goyette M.,
RA Graham L., Grand-Pierre N., Grant G., Hagos B., Heatford A., Horton L.,
RA Howland J.C., Iliev I., Johnson R., Jones C., Kann L., Karatas A.,
RA Klein J., Landers T., Larcocque K., Lehoczy J., Levine R., Lieu C.,
RA Liu G., Locke K., Macdonald P., Margulis N., McCarthy M., McEwan P.,
RA McGurk A., McKernan K., McPheters R., Meldrum J., Menus L., Mihova T.,
RA Miranda C., Mienna V., Morrow J., Naylor J., Norman C.H., O'Connor T.,
RA O'Donnell P., O'Neill D., Oliver T.M., Peterson K., Pierre N., Pisan C.,
RA Pollara V., Raymond C., Riley R., Rogov P., Rothman D., Roy A., Santos R.,
RA Schauer S., Severy P., Spencer B., Stange-Thoman N., Stojanovic N.,
RA Subramanian A., Talamas J., Testfave S., Theodore J., Tirrell A.,
RA Travers M., Trigilio J., Vassiliev H., Viel R., Vo A., Wilson B., Wu X.,
RA Wyman D., Ye W.J., Young G., Zainoun J., Zimmer A., Zody M.;
RT Submitted (28-FEB-2000) to the EMBL/GenBank/DBJ databases.
RL Whitehead Institute/MIT Center for Genome Research, 320 Charles Street,
RL Cambridge, MA 02141, USA
XX
XX On Mar 12, 2000 this sequence version replaced gi:7108199.
CC All repeats were identified using RepeatMasker:
CC Smit, A.F.A. & Green, P. (1996-1997)
CC <http://ftp.genome.washington.edu/RM/RepeatMasker.html>
CC ----- Genome Center -----
CC Center: Whitehead Institute/ MIT Center for Genome Research
CC Center code: WIBR
CC Web site: <http://www-seq.wi.mit.edu>
CC Contact: submissions@genome.wi.mit.edu
CC ----- Project Information -----
CC Center project name: L6834
CC Center clone name: 454.B.3
CC ----- Summary Statistics -----
CC Sequencing vector: M13; M7815; 100% of reads
CC Chemistry: Dye-terminator Big Dye; 100% of reads
CC Assembly program: Phrap; version 0.960731
CC Consensus quality: 157857 bases at least Q40
CC Consensus quality: 176423 bases at least Q30
CC Consensus quality: 186150 bases at least Q20
CC Insert size: 190000; agarose-fp
CC Insert size: 193118; sum-of-contigs
CC Quality coverage: 3.2 in Q20 bases; agarose-fp
CC Quality coverage: 3.2 in Q20 bases; sum-of-contigs
CC -----
CC * NOTE: This is a 'working draft' sequence. It currently
CC * consists of 44 contigs. The true order of the pieces
CC * is not known and their order in this sequence record is
CC * arbitrary. Gaps between the contigs are represented as
CC * runs of N, but the exact sizes of the gaps are unknown.
CC * This record will be updated with the finished sequence
CC * as soon as it is available and the accession number will
CC * be preserved.
CC 1 1100: contig of 1100 bp in length

```

CC * 1101 1200: gap of 100 bp
CC * 1201 2538: gap of 1238 bp in length
CC * 2439 2538: gap of 100 bp
CC * 2539 3781: contig of 1243 bp in length
CC * 3782 3881: gap of 100 bp
CC * 3882 5045: contig of 1164 bp in length
CC * 5046 5145: gap of 100 bp
CC * 5146 6642: contig of 1497 bp in length
CC * 6643 6742: gap of 100 bp
CC * 6743 8164: contig of 1422 bp in length
CC * 8165 8264: gap of 100 bp
CC * 8265 9726: contig of 1462 bp in length
CC * 9727 9826: gap of 100 bp
CC * 9827 11739: contig of 1913 bp in length
CC * 11740 11839: gap of 100 bp
CC * 11840 13330: contig of 1491 bp in length
CC * 13331 13430: gap of 100 bp
CC * 13431 15648: contig of 2219 bp in length
CC * 15650 15749: gap of 100 bp
CC * 15750 18374: contig of 2625 bp in length
CC * 18375 18474: gap of 100 bp
CC * 18475 20648: contig of 2174 bp in length
CC * 20649 20748: gap of 100 bp
CC * 20749 22015: contig of 1267 bp in length
CC * 22016 22115: gap of 100 bp
CC * 22116 24856: contig of 2741 bp in length
CC * 24857 24956: gap of 100 bp
CC * 24957 28230: contig of 3274 bp in length
CC * 28231 28330: gap of 100 bp
CC * 28331 30774: contig of 2444 bp in length
CC * 30775 30874: gap of 100 bp
CC * 30875 34408: contig of 3534 bp in length
CC * 34409 34508: gap of 100 bp
CC * 34509 36298: contig of 1790 bp in length
CC * 36299 36398: gap of 100 bp
CC * 36399 38662: contig of 2264 bp in length
CC * 38663 38762: gap of 100 bp
CC * 38763 42095: contig of 3333 bp in length
CC * 42096 42195: gap of 100 bp
CC * 42196 45055: contig of 2860 bp in length
CC * 45056 45155: gap of 100 bp
CC * 45156 47782: contig of 2627 bp in length
CC * 47783 47882: gap of 100 bp
CC * 47883 52304: contig of 4422 bp in length
CC * 52305 52404: gap of 100 bp
CC * 52405 55828: contig of 3424 bp in length
CC * 55829 55928: gap of 100 bp
CC * 55929 58658: contig of 3730 bp in length
CC * 58659 59758: gap of 100 bp
CC * 59759 63979: contig of 4221 bp in length
CC * 63980 64079: gap of 100 bp
CC * 64080 67509: contig of 3430 bp in length
CC * 67510 67609: gap of 100 bp
CC * 67610 72512: contig of 4903 bp in length
CC * 72513 72612: gap of 100 bp
CC * 72613 77741: contig of 5129 bp in length
CC * 77742 77841: gap of 100 bp
CC * 77842 83349: contig of 5508 bp in length
CC * 83350 83449: gap of 100 bp
CC * 83450 88963: contig of 5514 bp in length
CC * 88964 89063: gap of 100 bp
CC * 89064 93791: contig of 4728 bp in length
CC * 93792 93891: gap of 100 bp
CC * 93892 98121: contig of 4230 bp in length
CC * 98122 98221: gap of 100 bp
CC * 98222 103190: contig of 4969 bp in length
CC * 103191 103290: gap of 100 bp
CC * 103291 109512: contig of 6222 bp in length
CC * 109513 109612: gap of 100 bp
CC * 109613 115504: contig of 5892 bp in length
CC * 115505 115604: gap of 100 bp
CC * 115605 122119: contig of 5515 bp in length
CC * 121120 121219: gap of 100 bp

```

```

CC * 121220 125578: contig of 4359 bp in length
CC * 125579 125678: gap of 100 bp
CC * 125679 131585: contig of 5907 bp in length
CC * 131586 131685: gap of 100 bp
CC * 131686 138039: contig of 6354 bp in length
CC * 138040 138139: gap of 100 bp
CC * 138140 148214: contig of 10075 bp in length
CC * 148215 148314: gap of 100 bp
CC * 148315 158365: contig of 10051 bp in length
CC * 158366 158465: gap of 100 bp
CC * 158466 169691: contig of 11226 bp in length
CC * 169692 169791: gap of 100 bp
CC * 169792 197418: contig of 27627 bp in length.
XX
FH Key Location/Qualifiers
FH 1. 197418
FT source /chromosome="2"
FT /db_xref="taxon:9606"
FT /organism="Homo sapiens"
FT /map="2"
FT /clone="RP11-454B3"
FT /clone_lib="RPC1-11 Human Male BAC"
FT 1. 1100
FT /note="assembly_fragment"
FT 1201. 2438
FT misc_feature /note="assembly_fragment"
FT 2539. 3781
FT misc_feature /note="assembly_fragment"
FT 3882. 5045
FT misc_feature /note="assembly_fragment"
FT 5146. 6642
FT /note="assembly_fragment"
FT 6743. 8164
FT /note="assembly_fragment"
FT 8265. 9726
FT misc_feature /note="assembly_fragment"
FT 9827. 11739
FT misc_feature /note="assembly_fragment"
FT 11840. 13330
FT /note="assembly_fragment"
FT 13431. 15649
FT misc_feature /note="assembly_fragment"
FT 15750. 18374
FT /note="assembly_fragment"
FT
FT Query Match 42.9%; Score 27; DB 30; Length 197418;
Best Local Similarity 76.7%; Pred. No. 21;
Matches 33; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
Qy 3 ttccaagaagccagtcacatcatcactgcaacagagagact 45
Db 161203 TTCTAGCAATGACGATGCTATATCTATCTATCATGAGGACT 161161
|||||
RESULT 15
S75713/c 1632 bp mRNA linear ROD 09-AUG-2000
LOCUS S75713
DEFINITION abd-2:acetaldehyde dehydrogenase [mice, BALB/c, livers, mRNA
partial, 1632 nt].
ACCESSION S75713
VERSION S75713.1 GI:913809
KEYWORDS
SOURCE house mouse BALB/c livers.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Granulata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 1632)
AUTHORS Bond,S.L. and Singh,S.M.
TITLE DNA sequence analysis of the cytosolic acetaldehyde dehydrogenase
gene (Abd-2) in mouse strains with variable ethanol preferences
JOURNAL Biochem. Med. Metab. Biol. 52 (2), 155-159 (1994)
MEDLINE 95085815

```

REMARK

GenBank staff at the National Library of Medicine created this entry (NCBI gi|161480) from the original journal article. This sequence comes from Fig. 1A.

FEATURES

Source

Location/Qualifiers
1. .1632

```
/organism="Mus musculus"
/strain="BA1B/c"
```

```
/strain="BALB/c"
/db xref="taxon:10090"
```

```
/add_xref="axon:10090"
/tissue_type="liver"
```

1. .1632

gene	1. .1632
gene	/gene="ahd-2"
gene	1. 1500

CDS	1.1506
-----	--------

```
/gene="ahd-2"
/note="This s
```

translation presented here differs from translation in /note=... this sequence comes from Fig. 1A; conceptual

publication; A

```
/codon_start=1
```

```
/product="acetaldehyde  
/protein_id="AAP3275A.2"
```

```
/db_xref="GI:9755362"
```

/translation="MSSPAQPAVPAPLADLKIQHTKIFINNEWHNSVSGKKFPVLPNPA

SEEVICQVEEGDKADVKAVKARQAFQIGSPWRMDASERGLNKLADLMERDRLL
TAMERATVCCGVFANVATCETCCCTKATVCAQCAVDTTCOMTDCDCTEMVMBPBT

GYCGGOTIPWNEPMTETWKGSPALSCGNTVVYKPAFOTPI,TAI,HI,ASI,IKKAGTTPPGV
LATMEALNGKVFANAYLSDLGGCIKALKVCAGWADRIHGOTIPSDGDIETTYTRREPI

VNIVPGYCTAGAAISSHMDVDKVAFTGSTQVGKLIKEAGKSNLKRVTLELGKSPCC

IVFADADDIAVEFAHHGVYHQGCCVAASRIFVEESVYDEFVKRSVERAKKYVLGN

PLTPGINSQPDIDKEQHDKILDLIESGKKEGAKLECGGGRWGNKGFFVQPPVFSNVTLE
EUBTAVETETECEDUOCTMEKESYDQNTTPRANIMMUYCTABQIEMRDIKATMUSCAIQAQ

YVWVNCYMMI.SAOCPEGEKMGSGNGBEIGEHGILYEYTELKTVAAMKISOKNS."

341 c 426 g 405 t

BASE COUNT	460 a	341 c	426 g	405 t
ORIGIN				

Query Match	42.58;	Score 26.8;	DB 10;	Length 1632;
-------------	--------	-------------	--------	--------------

Matches 37; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 10 aagccagtgcacatcatctactgcacaagagagactggtaatgcccagagatg 63

Db 551 AGGCGAGGGCTATCTTCCAAATGACATGAGCATTGGAATAATCCAGGGATG 498

Search completed: May 12, 2002, 20:13:53
Job time: 7215 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 12, 2002, 19:41:35 ; Search time 2487.21 Seconds

(without alignments)
1069.029 Million cell updates/sec

Title: US-09-554-024-5

Perfect score: 197
Sequence: 1 ggcgtgcagcgcgtcgttc.....ggcgaacgtgcagcagatcc 197

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estmu:*
4: em_estnu:*
5: em_estrov:*
6: em_estrpl:*
7: em_estro:*
8: em_hic:*
9: gb_estl1:*
10: gb_estl2:*
11: gb_hic:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrlc:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	40.4	20.5	817	12	A2157865
2	39.8	20.2	1202	10	BE962592
3	39	19.8	266	12	A2493089
4	39	19.8	568	12	A2271424
5	38.6	19.6	1344	10	BC111381
6	38.2	19.4	266	12	BH124550
7	38.2	19.4	1163	10	BR310265
8	37.8	19.2	389	9	AA818956
9	37.8	19.2	997	12	CNS0057E
10	37.6	19.1	421	9	AA667914
11	37.6	19.1	440	12	A2170914
12	37.6	19.1	477	12	AC081531
13	37.6	19.1	492	9	BB853900
14	37.6	19.1	493	9	BB853873
15	37.6	19.1	769	12	A2206648
16	37.4	19.0	412	10	BR143006
17	37.4	19.0	416	10	BE283447

C 18	37.4	19.0	1303	10	BG026974
C 19	37.2	18.9	399	12	AQ821596
C 20	37.2	18.9	758	12	AG035520
C 21	36.8	18.7	320	9	AW198084
C 22	36.8	18.7	875	12	AG130893
C 23	36.8	18.7	1025	10	BG105107
C 24	36.8	18.7	1147	12	AQ895354
C 25	36.8	18.7	2099	10	BG294288
C 26	36.6	18.6	415	10	BF062823
C 27	36.6	18.6	965	12	BG823762
C 28	36.4	18.5	689	12	AG056509
C 29	36.4	18.5	1249	10	BF036764
C 30	36.4	18.5	1341	10	BF795965
C 31	36.2	18.4	302	10	W01197
C 32	36.2	18.4	979	12	CNS070C7
C 33	36.2	18.4	1571	10	BF101099
C 34	36	18.3	227	12	AG024885
C 35	36	18.3	348	10	BG535106
C 36	36	18.3	451	9	BB852401
C 37	36	18.3	465	9	BB854380
C 38	36	18.3	564	12	AZ630243
C 39	36	18.3	750	12	CNS01Y5V
C 40	36	18.3	967	10	BG535588
C 41	36	18.3	1001	12	AG107552
C 42	36	18.3	1024	10	BM415981
C 43	36	18.3	1219	10	BE966129
C 44	35.8	18.2	286	12	AQ103027
C 45	35.8	18.2	693	12	BH575352

ALIGNMENTS

RESULT 1
LOCUS A2157865
DEFINITION SP_0059_A1-B04-SB6E Strongylocentrotus purpuratus, purple sea urchin, sperm genomic BAC library Strongylocentrotus purpuratus genomic clone Plate=59 Col=7 Row=C, DNA sequence.
ACCESSION A2157865
VERSION A2157865.1 GI:8310468
KEYWORDS GSS.
SOURCE Strongylocentrotus purpuratus.
ORGANISM Strongylocentrotus purpuratus.
REFERENCE Cameron,R.A., Mahalras,G., Rast,J.P., Martinez,P., Biondi,T.R., Swartzell,S., Wallace,J.C., Poustka,A.D., Livingston,B.T., Wray,G.A., Ettensohn,C.A., Lehnach,H., Britten,R.J., Davidson,E.H. and Hood,L.
TITLE A sea urchin genome project: Sequence scan, virtual map, and additional resources
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 97 (17), 9514-9518 (2000)
MEDLINE 20402566
COMMENT Contact: Cameron, RA, Davidson, EH, Hood, L
Division of Biology 156-29
California Institute of Technology
Pasadena California 91125, USA
Tel: (626) 395-8421
Fax: (626) 793-3047
Email: acameron@caltech.edu
Plate: 59 row: C column: 7
Seq primer: SP6
Class: BAC ends
High quality sequence stop: 817.
Location/Qualifiers
1. 817
/organism="Strongylocentrotus purpuratus"
/db_xref="taxon:7668"
/clone="plate=59 Col=7 Row=C"
/clone_lib="Strongylocentrotus purpuratus, purple sea

```

urcln, sperm genomic BAC library"
/notes="Organ: sperm; Vector: BACs.6; BAC Clones in E-Coli:
DH10B"
BASE COUNT      153 a      217 c      141 g      306 t
ORIGIN

```

[illegible]

RESULT	2
BE962592/c	
LOCUS	1202 bp mRNA linear EST 14-DEC-2000
DEFINITION	B01655933R1 NIH_MGC_66 Homo sapiens CDNA clone IMAGE:3855729 3,
ACCESSION	B01655933R1
VERSION	mRNA sequence.
KEYWORDS	BE962592
SOURCE	human.

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Eukaryota: Metazoa; Chordata; Craniata; Vertebrata; Euteleostei
Mammalia; Euthalia; Primates; Carnivora; Homiidae; Homo.
1 (bases 1 to 1202)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
On Oct 3, 2000 this sequence version replaced gi:10573297.

Email: c9apbs-r@mail.nih.gov
 Tissue Procurement: DCTD/DTF
 cDNA library preparation: Life Technologies, Inc.
 cDNA library arrayed by: the I.M.A.G.E. Consortium (NLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/NLNL at:
<http://image.llnl.gov>
 Plate: LLCMS70 row: e column: 10
 High quality sequence stop: 55

FEATURES
SOURCE

```

1. .1202
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="3855729"
/clone_id="NH.MCC.66"
/clone_type="adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
note="Organ: ovary; Vector: pCMV-Sport6; Site:1: NotI; Site:2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.8 kb. Library constructed by Life Technologies."

```

BASE COUNT	654 a	101 c	386 g	61 t
ORIGIN				

Query Match	20.2%;	Score 39.8;	DB 10;	Length 1202;
Best Local Similarity	64.8%;	Pred. NO. 42;		
Matches 59; Conservative	0;	Mismatches 32;	Indels 0;	Gaps 0;

Oy 17 ttctgtagctttcttcctcagctccacatcttccacttgtagtcaactctctctcttctct 76
||| | | | | | | | | | | | | | | | |
Db 979 tttctttcccttttgattttcccttctcttctctctctctctctctctcttcttctt 920

oy 77 tgtgattctctcaactctctgacgtgacggttcc 107

Db 919 TCTCCTCTTTTCCCTTGCTCTCCCTCTCC 889

RESULT	3
AZ493089/c	
LOCUS	266 bp DNA linear GSS 05-OCT-2000
DEFINITION	Mouse 10kb plasmid UUCGIM3 library Mus musculus genomic clone UUCGIM327M20 F, DNA sequence.
ACCESSION	AZ493089
VERSION	AZ493089
KEYWORDS	AZ493089.1 GI:1066263
SOURCE	GSS.
ORGANISM	house mouse. Mus musculus

REFERENCE
1 (bases 1 to 266)
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamill, C., ...

TITLE	Mouse whole genome scaffolding with paired end reads from 10xb
JOURNAL	plasmid inserts
COMMENT	Unpublished (2000)
	Contact: Robert B. Weiss

Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLc, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0327 row: M column: 20
Seq primer: CGTGTGTAACGACGGCCACT
Class: plasmid ends
High quality sequence stop: 266.

FEATURES
source

```
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="U06C1M0327M20"
/clone_lib="Mouse 10kb plasmid U06C1M library"
```

1a6_host="E. coli strain XL10-gold, T1-resistant, F⁻/note="Vector: pMD24::purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (<http://www.jax.org/resources/documents/dnars/>). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-treated with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD24 (g114732114[gb|AF128072.1]), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptor complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT	107 a	31 c	86 g	42 t
ORIGIN				

Query Match	19.8%;	Score 39;	DB 12;	Length 266;
Best Local Similarity	59.5%;	Pred. No. 65;		
Matches 66;	Conservative 0;	Mismatches 45;	Indels 0;	Gaps 0;

[illegible]

Db	490	TTCCTTCTTTCTGTCGTCCATCATCATGTAGCCACAGACCAGCCTTGACT	440
RESULT	5		
BGI11381/c			
LOCUS			
DEFINITION	BGI11381	1344 bp	mRNA linear EST 30-JAN-2001
ACCESSION	602281745f1 NIH_MGC_86 Homo sapiens cDNA clone IMAGE:4369259 5'		
VERSION	BGI11381		
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE	1 (bases 1 to 1344)		
JOURNAL	NIH-MGC http://mgc.nci.nih.gov/		
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgapbs-r@mail.nih.gov Tissue Procurement: ATCC CDNA Library Preparation: Life Technologies, Inc. CDNA library Arrayed by: Incyte Genomics, Inc. DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LHAM10024 row: j column: 12 High quality sequence stop: 520. Location/Qualifiers 1..1344 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:4369259" /clone.lib="NIH_MGC_86" /tissue.type="osteosarcoma, cell line" /lab.host="RDH10 (phage-resistant)" /note="Organ: bone, Vector: pCMV-Sport6; Site:1; Nctti: Site:2; Salt: Cloned unidirectionally; oligo-dt primed. Average insert size 1.33 Kb. Library enriched for full-length clones and constructed by life technologies. Note: this is a NIH_MGC Library."		
BASE COUNT	446 a 396 c 393 g 109 t		
ORIGIN			
Query Match	19.6% Score 38.6; DB 10; Length 1344;		
Best Local Similarity	53.7%; Pred. No. 76;		
Matches	80; Conservative 0; Mismatches 69; Indels 0; Gaps 0;		
OY	18 ttctgtctttcttcacgtccatcttccttccttggtctactcttcttccct 77		
Db	219 TACCTGGTGTGCTTCGCCCTTCCTCCCTCTCTCTCACTTCCTCTTCT 160		
OY	78 gtgattcttaacctgcgcggtgccagaagaccagtccaatcatctactgaac 137		
Db	159 CTTCCTTCCTCTCTTCCTTGTACTTCGCCCGCTTCCTCCGACATCATCTGAGAGAAG 100		
OY	138 agagagaaactgttagtgcagagaggtgtg 166		
Db	99 AGGAGAGGACGACGAGGAAGAAGACGAG 71		
RESULT	6		
BHI124550			
LOCUS			
DEFINITION	BHI124550	266 bp	DNA linear GSS 19-JUL-2001
ACCESSION	RPCI-24-289F3.TV RPCI-24 Mus musculus genomic clone RPCI-24-289F3,		
VERSION	BHI124550		
KEYWORDS	GSS.		

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT	
Mammalia: Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akhret,B., Levins,M., Tsagarev,G., Geer,K., Krol,M., Shvarshbeyn,A., Gebregeorgis,E., Russell,D., de Jong,P. and Fraser,C.M	Mouse BAC End Sequences from Library RPCI-24	Other_GSSS: RPCI-24-289F3.TJ	Contact: Shaying Zhao Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 0200 Fax: 301 838 0208 Email: szhao@tigr.org	
house mouse.				Clones are derived from the mouse BAC library RPCI-24. For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources (http://www.choil.org/bacpac/orderingframe.htm). BAC end plate: 289 tww: F column: 3	
Nus musculus				Seq primer: T2	
Class: BAC ends.					
FEATURES	Source	Location/Qualifiers			
		1..266			
		/organism="Mus musculus"			
		/strain="C57BL/6J"			
		/db_xref="taxon:10090"			
		/clone="RPCI-24-289F3"			
		/clone_1lb="RPCI-24"			
		/sex="Male"			
		/cell_type="Spleen/Brain"			
		/note="Vector: pTRABAC1. Site_1: BamHI; Site_2: BamHI; RPCI-24 Mouse BAC library produced by Pieter de Jong. The library was cloned in the pTRABAC1 cloning vector at the BamI sites using MboI partially digested male C57BL/6J DNA."			
BASE COUNT	47 a	87 c	18 g	114 t	
ORIGIN					
Query Match	19.4%	Score 38.2;	DB 12;	Length 266;	
Best Local Similarity	73.1%;	Pred. NO. 96;			
Matches 49;	Conservative	0;	Mismatches 18;	Indels 0;	Gaps 0;
Oy	26	tcttcctcagactccatcttcctctctgtgcbctcatctctcttcttcttgtagctc	85		
	1		1		1
	1		1		1
	1		1		1
	1		1		1
D -	171	TCCTCTCTCTCTCCCTTCCCTCTCTCTCTTTTTCCTCTCTCTCTGTCATCATGC	230		
Oy	86	tcaacct 92			
	1				
	1				
b	231	TCCTCTCT 237			
RESULT	7				
BF310265/c					
LOCUS	BF310265	1163 bp	mRNA	linear	EST 21-NOV-2000
DEFINITION	601894769F1 NIH_MGC_19 Homo sapiens CDNA clone IMAGE:4124249 5'				
ACCESSION	BF310265				
VERSION	BF310265.1	GI:11257806			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
REFERENCE	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	NIH-MGC http://mgc.ncl.nih.gov/ .				
TITLE	1 (bases 1 to 1163)				
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)				
	Unpublished (1999)				

COMMENT

Contact: Robert Strausberg, Ph.D.
Email: c9apbs.r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: Image.llnl.gov
Place: LBL010 row: 1 column: 18
High quality sequence stop: 506.

FEATURES

Source

Location/Qualifiers

1..1163
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4124249"
/clone_1lb="NIH-MGC-19"
/tissue_type="neuroblastoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: brain; Vector: pORF7; Site_1: XhoI; Site_2: EcoRI; CDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GCGACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald W. Rubin (University of California, Berkeley) using Zap-CDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: This is a NIH-MGC Library."

BASE COUNT

325 a 297 c 348 g 192 t Others

ORIGIN

Query Match 19.4%; Score 38.2; DB 10; Length 1163;
Best Local Similarity 63.7%; Pred. No. 93;
Matches 58; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

OY 9 gccacgagctgcgttcctcctccagctccatccctctcctcttgtagctactctctt 68
|| | ||| ||||| || | ||| ||||| || ||||| || |||
Db 1058 GCATCGCTTCCTTGCTTCTCTTGGTTTCCTTCCTTCGCGCCTGTGCTCCCTT 999
||| || | || ||||| || | ||||| |||

OY 69 ctcttccttgatctctcacctctcgscgtg 99
||| || | || ||||| || | ||||| |||

Db 998 CTTCGCTCTGCT 968

RESULT

8

AAB18956

LOCUS AAB18956 389 bp mRNA linear EST 03-JUL-1999

DEFINITION U1-R-A0-as-e-11-0-U1.s1 U1-R-A0 Rattus norvegicus CDNA clone

UI-R-A0-as-e-11-0-U1.3 similar to gb|AF001308|ATAF001308 Arabidopsis thaliana chromosome IV BAC clone T10M13, complete sequence [Arabidopsis thaliana], mRNA sequence.

ACCESSION

AAB18956

VERSION

AAB18956.1 GI:4227624

KEYWORDS

EST.

SOURCE

Norway rat.

ORGANISM

Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathu; Muridae; Murinae;
Rattus.

REFERENCE

1 (bases 1 to 389)
Bonaldi,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene discovery

JOURNAL MEDLINE COMMENT

Genome Res. 6 (9), 791-806 (1996)
97044477
On Feb 17, 1998 this sequence version replaced gi:2880542.
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msotares@blue.weeg.uiowa.edu
The sequence tag present in the CDNA between the NotI site and the

oligo-dr track served to identify it as a clone from the normalized adult Brain library. cDNA library Preparation: M. Fatima Bonaldo, Ph.D. Clone distribution: clones will be available through Research Genetics. This clone is also available through the I.M.A.G.E. Consortium at LNL (info@image.lnl.gov). IMAGE ID=167986 The following repetitive elements were found in this cDNA sequence: 1-92, >POLY-A#simple_repeat 310-389, >B2#SINE/B2 Seq primer: M13 Forward

FEATURES

source
Location/Qualifiers
1. 389
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UT-R-A0-as-e-11-0-UT"
/clone_lib="UT-R-A0"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/note="vector: pTZ19-pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; This library consists of a mixture of individually tagged normalized libraries constructed from rat placenta, adult lung, brain, liver, kidney, heart, spleen, ovary, and muscle. The tag is a string of 3-5 nucleotides present between the Not I site and the oligo-dr track which allows identification of the library of origin of a clone within the mixture."

BASE COUNT
86 a 82 c 74 g 147 t

ORIGIN

Query Match
Best Local Similarity 53.8%; Score 37.8; DB 9; Length 389;
Matches 78; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

Oy 17 ttctgctctctctcagctctcctctctctctctctctctctctct 76
|||||
Db 29 ttct 88
|||||
Oy 77 tctgctctctcctctctctctctctctctctctctctctct 136
|||||
Db 89 tcttccaccatcatcatcatcatcatcatcatcatcatcatcat 148
|||||

Oy 137 caggagagctgtaagtcagag 161
|||||
Db 149 CAAGAGAGTCTCGATCCGAG 173
|||||

RESULT 9
CNS005TE 997 bp DNA linear GSS 03-JUN-1999
LOCUS
DEFINITION
CNS005TE 997 bp DNA linear GSS 03-JUN-1999
Drosophila melanogaster genome survey sequence JET3 end of BAC #
BACR1K22 of RPL-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
AL060767.1 GI:4943573
GSS.
fruit fly.
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE
AUTHORS
TITLE
JOURNAL
Genoscope.
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see <http://www.fruitfly.org> The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osogawa and

Aaron Mammoser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPL-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp, the same strain used for the BDGP's
p1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

source
Location/Qualifiers
1. 997
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPL-98"
/clone="BACR1K22"
/note="end : TET3"
TET3
BASE COUNT
89 a 99 c 13 g 238 t 538 others

ORIGIN

Query Match
Best Local Similarity 36.4%; Score 37.8; DB 12; Length 997;
Matches 28; Conservative 33; Mismatches 16; Indels 0; Gaps 0;

Oy 17 ttctgctctctctcagctctcctctctctctctctctctctct 76
|||||
Db 818 tcttctctctctctctctctctctctctctctctctctctct 877
|||||

Oy 77 tctgctctctcctctct 93
|||||
Db 878 tcttctctctctctctctctctctctctctctctctctct 894
|||||

RESULT 10
AA667914 421 bp mRNA linear EST 19-NOV-1997
LOCUS
DEFINITION
AA667914 421 bp mRNA linear EST 19-NOV-1997
IMAGE:1222889 5', mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
AA667914.1 GI:2626615
EST.
house mouse.
Mus musculus

REFERENCE
AUTHORS
Mammoser, A., Allen, M., Bowles, M., Dietrich, N., Dubouque, T.,
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Stepien, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.

COMMENT
The WashU-HMI Mouse EST Project
Unpublished (1996)
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu

High quality sequence stop: 399.
Seq primer: -28mJ rev1 ET from Amersham
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
WGI:648481
Location/Qualifiers
1. 421
/organism="Mus musculus"
/strain="NIH/Swiss"
/db_xref="taxon:10090"
/clone="IMAGE:1222889"
/clone_lib="Stratagene mouse heart (#937316)"
/sex="pooled"

```
/dev_stage="13 day embryos"
/lab_host="SOLR (kanamycin resistant)"
/notes="Organ: heart; Vector: pBluescript SK-; Site_1:
EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
Oligo dt. 93 pooled NIH/Swiss 13 day embryo hearts.
Average insert size: 1.0 kb; Uni-ZAP XR Vector: -5'
adaptor sequence: 5' GAATTCGACGAG 3' -3' adaptor
sequence: 5' CTCGAGTTTCTTTTCTTTTCTTTT 3'"

BASE COUNT      70 a      139 c      67 g      145 t
ORIGIN

Query Match      19.1%; Score 37.6; DB 9; Length 421;
Best Local Similarity 59.3%; Pred. No. 1.3e+02;
Matches 64; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

OY 1 gcctgcagccatgggttcgtcttccttcacagctccatcttccttcgtccta 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 195 GGTCTTTGTTTGTATTTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTA 254
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 61 cctctctctctctctctctctctctctctctctctctctctctctctca 108
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 255 CCCATCTTACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCA 302
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 11
LOCUS      A2170914      440 bp      DNA      linear      GSS 29-AUG-2000
DEFINITION SP_0118_1_B02_77A Strongylocentrotus purpuratus, purple sea urchin
            clone Plate=118 Col=3 Row=C, DNA sequence.
ACCESSION   A2170914
VERSION     A2170914.1 GI:8341282
KEYWORDS    GSS.
SOURCE      Strongylocentrotus purpuratus.
            Eukaryote; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
            Echinidea; Euechinoidea; Echinacea; Echinoidea;
            Strongylocentrotidae; Strongylocentrotus.
REFERENCE   1 (bases 1 to 440)
            Cameron,R.A., Mahairas,G., Rast,J.P., Martinez,P., Biondi,T.R.,
            Swartzell,S., Wallace,J.C., Pouška,A.J., Livingston,B.T., Wray,
            G.A., Ettensohn,C.A., Lehrach,H., Britten,R.J., Davidson,E.H. and
            Hood,L.
            A sea urchin genome project: Sequence scan, virtual map, and
            additional resources
JOURNAL     Proc. Natl. Acad. Sci. U. S. A. 97 (17), 9514-9518 (2000)
MEDLINE     20402565
COMMENT     Contact: Cameron, RA, Davidson, EH, Hood, L
            Division of Biology 156-29
            California Institute of Technology
            Pasadena California 91125, USA
            Tel: (626) 395-8421
            Fax: (626) 793-3047
            Email: acameron@caltech.edu
            Plate: 118 row: C column: 3
            Seq primer: 77
            Class: BAC ends
            High quality sequence stop: 440.
            Location/Qualifiers
                1..440
                /organism="Strongylocentrotus purpuratus"
                /db_xref="taxon:7668"
                /clone="Plate=118 Col=3 Row=C"
                /clone_1db="Strongylocentrotus purpuratus, purple sea
                urchin, sperm genomic BAC library"
                /note="Organ: sperm; Vector: BAC3.6; BAC Clones in E-Coli
                DH10B"

BASE COUNT      85 a      124 c      60 g      168 t      3 others
ORIGIN
```

```
Query Match      19.1%; Score 37.6; DB 12; Length 440;
Best Local Similarity 72.1%; Pred. No. 1.3e+02;
Matches 49; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

OY 10 ccattggttcgtctctctctctccatctctctctctctctctctctc 69
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 315 CCGTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCC 374
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 70 ttctctct 77
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 375 TTTCCTT 382

RESULT 12
LOCUS      AO881531/C      477 bp      DNA      linear      GSS 09-NOV-1999
DEFINITION HS_5253_1_F03_77A RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate=9021 Col=5 Row=K, DNA sequence.
ACCESSION   AO881531
VERSION     AO881531.1 GI:6312998
KEYWORDS    GSS.
SOURCE      human.
            Homo sapiens
            Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
            1 (bases 1 to 477)
            Mahairas,G.G., Wallace,J.C., Smith,R., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
            Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
JOURNAL     Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE     99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3687
            Email: jwallace@u.washington.edu
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
            or from Resear h Genetics (info@resgen.com). BAC end Web Server:
            http://www.htsc.washington.edu
            Plate: 9021 row: K column: 5
            Seq primer: 77
            Class: BAC ends
            High quality sequence stop: 477.
            Location/Qualifiers
                1..477
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="Plate=9021 Col=5 Row=K"
                /clone_1db="RPCI-11 Human Male BAC Library"
                /sex="male"
                /note="Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI;
                Male blood DNA was isolated from one randomly chosen donor
                and partially digested with a combination of EcoRI and
                EcoRI methylase. Size selected DNA was cloned into the
                pBAC3.6 vector at EcoRI sites"

BASE COUNT      273 a      8 c      115 g      26 t      55 others
ORIGIN

Query Match      19.1%; Score 37.6; DB 12; Length 477;
Best Local Similarity 56.8%; Pred. No. 1.3e+02;
Matches 67; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

OY 17 ttctgcctctctctcagctccatctctctctctctctctctctctctcc 76
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 297 TTTCCTTATTTCTCTTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTTTC 238
```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 12, 2002, 21:35:47 ; Search time 96.6 Seconds
(without alignments)
799.085 Million cell updates/sec

Title: US-09-554-024-5

Perfect score: 197
Sequence: 1 ggcgcacgcacatgggttc.....gcgaacgtgtcgcacgatcc 197

Scoring table: IDENTITY_NDC
Gapop 10.0 , Gapext 1.0

Searched: 398110 seqs, 195917913 residues

Total number of hits satisfying chosen parameters: 796220

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Pending_Patents_NA_New : *
1: /cgn2_6/ptodata/1/pna/PCY_NEW_COMB.seq : *
2: /cgn2_6/ptodata/1/pna/US06_NEW_COMB.seq : *
3: /cgn2_6/ptodata/1/pna/US07_NEW_COMB.seq : *
4: /cgn2_6/ptodata/1/pna/US08_NEW_COMB.seq : *
5: /cgn2_6/ptodata/1/pna/US09_NEW_COMB.seq : *
6: /cgn2_6/ptodata/1/pna/US10_NEW_COMB.seq : *
7: /cgn2_6/ptodata/1/pna/US60_NEW_COMB.seq : *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	34	17.3	42999	6	US-10-125-767-17
2	33.2	16.9	1377	6	US-10-105-299-1441
3	33.2	16.9	4182	5	US-09-782-959A-17
4	33.2	16.9	5754	1	PCT-US02-09785-1053
5	33.2	16.9	5754	6	US-10-105-299-13322
6	33.2	16.9	5755	1	PCT-US02-09785-1056
7	33.2	16.9	5755	6	US-10-105-299-13331
8	33	16.8	626	1	PCT-US02-01109-32
9	33	16.8	1199	6	US-10-114-893-128
10	32.2	16.3	6807	6	US-10-105-299-10493
11	31.4	15.9	9718	1	PCT-US02-09105-423
12	31.4	15.9	9718	1	PCT-US02-09188-882
13	31.4	15.9	9718	1	PCT-US02-09370-909
14	31.4	15.9	9718	6	US-10-105-299-6278
15	31.4	15.9	9728	1	PCT-US02-09105-422
16	31.4	15.9	9728	1	PCT-US02-09188-881
17	31.4	15.9	9728	1	PCT-US02-09370-908
18	31.4	15.9	9728	6	US-10-105-299-6277
19	31.4	15.9	180557	6	US-10-003-806-6
20	31.4	15.9	180557	6	US-10-003-806-9
21	31.2	15.8	1287	1	PCT-US02-06415-38
22	31.2	15.8	2593	1	PCT-US02-09785-1117
23	31.2	15.8	39119	1	PCT-US02-09188-1144
24	31.2	15.8	39119	1	PCT-US02-09257-718
25	31.2	15.8	39119	1	PCT-US02-09370-1181
26	31.2	15.8	39119	1	PCT-US02-09922-750

C	27	31.2	15.8	39119	6	US-10-105-299-6650	Sequence 6650, Ap
C	28	31.2	15.8	72604	5	US-09-691-064-7	Sequence 7, Appl
C	29	31	15.7	544	6	US-10-126-103-38	Sequence 38, Appl
C	30	31	15.7	43391	6	US-10-105-299-14118	Sequence 14118, A
C	31	30.6	15.5	771	6	US-10-116-051-1	Sequence 1, Appl
C	32	30.2	15.3	51402	6	US-10-105-299-12126	Sequence 12126, A
C	33	30	15.2	519	5	US-09-539-331D-31789	Sequence 31789, A
C	34	29.8	15.1	220	5	US-09-975-254-17338	Sequence 17338, A
C	35	29.8	15.1	242	5	US-09-975-254-11819	Sequence 11819, A
C	36	29.6	15.0	4893	6	US-10-000-256A-27	Sequence 27, Appl
C	37	29.6	15.0	39380	6	US-10-105-299-12733	Sequence 12733, A
C	38	29.6	14.9	1014	6	US-09-863-776-25	Sequence 25, Appl
C	39	29.4	14.9	80374	6	US-10-041-018-376	Sequence 376, App
C	40	29	14.7	1490	6	US-10-105-299-1403	Sequence 1403, App
C	41	29	14.7	1722	7	US-60-365-384-161	Sequence 161, App
C	42	29	14.7	2041	6	US-10-105-299-1739	Sequence 1739, App
C	43	29	14.7	2054	6	US-10-105-299-1740	Sequence 1740, Ap
C	44	29	14.7	2056	6	US-10-105-299-10538	Sequence 10538, A
C	45	29	14.7	2056	6	US-10-105-299-10540	Sequence 10540, A

ALIGNMENTS

RESULT 1
US-10-125-767-17
Sequence 17, Application US/10125767
GENERAL INFORMATION:
APPLICANT: Hadlaczky, Gyula
TITLE OF INVENTION: ARTIFICIAL CHROMOSOMES, USES THEREOF AND METHODS
FOR PREPARING ARTIFICIAL CHROMOSOMES
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:
ADDRESSEE: Heller Ehrman White & McCauliffe LLP
STREET: 4350 La Jolla Village Drive, 7th Floor
CITY: San Diego
STATE: CA
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/125,767
FILING DATE: 17-Apr-2002
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/724,693
FILING DATE: 28-Nov-2000
APPLICATION NUMBER: 08/835,682
FILING DATE: 10-Apr-1997
APPLICATION NUMBER: 08/695,191
FILING DATE: 07-Aug-1996
APPLICATION NUMBER: 08/682,080
FILING DATE: 15-Jul-1996
APPLICATION NUMBER: 08/629,822
FILING DATE: 10-Apr-1996
ATTORNEY/AGENT INFORMATION:
NAME: Seidman, Stephanie L.
REGISTRATION NUMBER: 33,779
REFERENCE/DOCKET NUMBER: 24601-4021
TELECOMMUNICATION INFORMATION:
TELEPHONE: 858-450-8403
TELEFAX: 858-587-5360
TELEX: <unknown>
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 42899 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

[illegible]

```

1  RESULT 6
2  PCT-US02-09785-1058/c
3  Sequence 1058, Application PC/TUS0209785
4  GENERAL INFORMATION:
5  APPLICANT: Human Genome Sciences, Inc.
6  TITLE OF INVENTION: Human Secreted Proteins
7  FILE REFERENCE: PS905PC
8  CURRENT APPLICATION NUMBER: PCT/US02/09785
9  CURRENT FILING DATE: 2002-03-19
10 PRIOR APPLICATION NUMBER: US 60/331,287
11 PRIOR FILING DATE: 2001-11-13
12 PRIOR APPLICATION NUMBER: US 60/306,171
13 PRIOR FILING DATE: 2001-07-19
14 PRIOR APPLICATION NUMBER: US 60/277,340
15 PRIOR FILING DATE: 2001-03-21
16 NUMBER OF SEQ ID NOS: 1130
17 SOFTWARE: Patentin Ver. 2.0
18 SEQ ID NO 1058
19 LENGTH: 5755
20 TYPE: DNA
21 ORGANISM: Homo sapiens
22 PCT-US02-09785-1058

```

Query Match	16.9%	Score 33.2	DB 1	Length 5755
Best Local Similarity	67.1%	Pred. No. 5.67		
Matches	47	Conservative	0	Mismatches 23
				Indels 0
				Gaps 0
QY	26	ttctctcagcttcacattctccctctgctgctcctctctctctctctcttgatc	85	
Db	2730	ttttcttttcccttttcttctctctcttcttcttcttcttcttcttcttcttctc	2671	
QY	86	tcaactctgc	95	
Db	2670	tcctctcttctc	2661	

```

RESULT 7
US-10-105-299-13331/C
: Sequence 13331. Application US/10105299
: GENERAL INFORMATION:
: APPLICANT: Rosen, et. al
: TITLE OR INVENTION: Human Secreted Proteins
: FILE REFERENCE: P5950
: CURRENT APPLICATION NUMBER: US/10/105,299
: CURRENT FILING DATE: 2002-03-26
: NUMBER OF SEQ ID NOS: 15197
: Prior Application removed - See File Wrapper or Paim
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 13331
:
: LENGTH: 5755
:
: TYPE: DNA
:
: ORGANISM: Homo sapiens
US-10-105-299-13331

```

	Query Match	16.9%	Score 33.2;	DB 6;	Length 5755;	
	Best Local Similarity	67.1%;	Pred. No. 5.6;			
	Matches 47;	Conservative	0;	Mismatches 23;	Indels 0;	Gaps 0.
OY	26	tctctcgaagcttcacatcttccctttgtgatactactcttcttcttcttgtagarc	85			

Dd 2730 TTTCTTCCTTTTCTTCCTCCTTCCTTCCTTCCTTCCTTCCTTC
QY 86 tcaactcttgc 95
TTTTTTTTT
Dd 2670 TCCTCTCTTC 2661

```

RESULT 8
PCT-US02-01109-32
; Sequence 32. Application PC/TUS0201109
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: 50 Human Secreted Proteins
; FILE REFERENCE: P2016PC12
; CURRENT APPLICATION NUMBER: PCT/US02/01109
; CURRENT FILING DATE: 2002-01-17
; PRIOR APPLICATION NUMBER: US 60/262,066
; PRIOR FILING DATE: 2001-01-18
; NUMBER OF SEQ. ID NOS: 206
; SOFTWARE: PatentIn Ver. 2.0
; SEQ. ID NO 32
; LENGTH: 626
; TYPE: DNA
; ORGANISM: Homo sapiens
PCT-US02-01109-32

```

Query Match	16.8%	Score 33	DB 1	Length 626
Best Local Similarity	60.7%	Pred. No. 4.3		
Matches 54	Conservative	0	Mismatches 35	Indels 0
Gaps				0
QY 17	tttcgagcttctctcgaagcttcacatcttcctctctggtcctacatctctcttctct	76		
DB 33	ttctctctcttcttctccctctacctcttcttggtggtgtgtcttctctctctctcc	92		
QY 77	tgtatctctcaactcttgccggtgcggtt	105		
DB 93	tttgagatttttttgtgtgtgttctctt	121		

```

RESULT      9
US-10-114-893-128
; Sequence 128, Application US/10114893
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M.
; APPLICANT: Lavallie, Edward R.
; APPLICANT: Collins-Racie, Lisa A.
; APPLICANT: Evans, Cheryl
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Bowman, Michael R.
; APPLICANT: Spaulding, Vikki
; APPLICANT: Carlin-Duckett, McKeough
; APPLICANT: Kelleher, Kerry S.
; APPLICANT: Genetics Institute, Inc.
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
; FILE REFERENCE: GI 6000-10A
; CURRENT APPLICATION NUMBER: US/10/114,893
; CURRENT FILING DATE: 2002-04-02
; EARLIER APPLICATION NUMBER: 09/413,232
; EARLIER FILING DATE: 1999-10-06
; NUMBER OF SEQ ID NOS: 321
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 128
; LENGTH: 1199
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-114-893-128

```

Query Match	16.8%;	Score 33;	DB 6;	Length 1199;
Best Local Similarity	60.7%;	Pred. No. 4.8;		

: PRIOR APPLICATION NUMBER: US 09/950,083
 : PRIOR FILING DATE: 2001-09-12
 : NUMBER OF SEQ. ID NOS: 1834
 : SOFTWARE: PatentIn Ver. 2.0
 : SEQ. ID NO 909
 : LENGTH: 9718
 : TYPE: DNA
 : ORGANISM: Homo sapiens
 : CCT-0502-09370-909

Query Match	15.9%	Score 31.4;	DB 1;	Length 9718;
Best Local Similarity	48.1%	Pred. No. 18;		
Matches	89;	Mismatches	96;	Gaps 0;

Oy	4	tcgaagccatctggattcttcgagctttctccaaagttccaaatcttcctcttctgtctcaac	63
Db	7075	tcctccctctctggatgacgtggaactgcaggaacgtgcacacatgcctgctggtttttatt	7016
Oy	64	ttctctcttctcttgagacttcaactcttgcgcgtgttccaaagaacgaatgcga	123
Db	7015	tttttgattttttgacagctctcactatgtttggccacgactgctcaaacatctttagcctcc	6956
Oy	124	tcaatcaactgaacaagaggaactgtaagtccaaagaaatgtagctcggcagagcgaa	183
Db	6955	accaatcttcccgccctcacacttccctgaaagctgtaagtctccggatattacagggccact	6896

QY	184	gtgtc	188
Db	6895	GTGCC	6891

```

RESULT 14
US-10-105-299-6278/c
: Sequence 6278, Application US/10105299
: GENERAL INFORMATION:
: APPLICANT: Rosen, et. al
: TITLE OF INVENTION: Human Secreted Proteins
: FILE REFERENCE: P9350
: CURRENT APPLICATION NUMBER: US/10/105,299
: CURRENT FILING DATE: 2002-03-26
: NUMBER OF SEQ ID NOS: 15197
: Prior Application removed - See File Wrapper or Palm
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 6278
: LENGTH: 9718
: TYPE: DNA
: ORGANISM: Homo sapiens
US-10-105-299-6278

```

Query Match	15.9%;	Score 31.4;	DB 6;	Length 9718;
Best Local Similarity	48.1%;	Pred. No. 18;		
Matches	89;	Conservative	0;	Mismatches 96;
				Indels 0;
				Gaps 0

Oy	4	tcgacgcacatggattcttcgcttcctttctctcaagcttccatcttccctctcttctgctctac	63
Db	7075	TCCTCCTCCGTGGGTAGCTGGAGCTGAGCAGCAGCAGCCACCATGCTCGCTGCTTTTATT	7016
Oy	64	tctctcttctctcttgatcctcctcactctctgcgctgcttccaaagaagcagctggcaa	123
Db	7015	TTTTTAACTTTTGTAGAGGCTCACTCACTAGTTGCCAGGCTGCTTCCAAACTTTGGGCTCC	6955
Oy	124	tcatctactgcaacagagagactctgtaagtgcacagagatgtlgagctcgcgcagagcgaa	183
Db	6955	ACCAATCTTCCGCCCTCAGGCTTCTGTAGAGCTGTAAAGTGTGGGAAATTACAGGTCCTACT	6896

QY	184	gtgtc	188
Db	6895	GTGCC	6891

RESULT 15

PCT-US02-09105-422/c
 : Sequence 422, Application PC/TUS0209105
 : GENERAL INFORMATION:
 : APPLICANT: Human Genome Sciences, Inc.
 : TITLE OF INVENTION: Human Secreted Proteins
 : FILE REFERENCE: PS951PCT
 : CURRENT APPLICATION NUMBER: PCT/US02/09105
 : CURRENT FILING DATE: 2002-03-26
 : PRIOR APPLICATION NUMBER: US 60/2278,650
 : PRIOR FILING DATE: 2001-03-27
 : PRIOR APPLICATION NUMBER: US 09/950,082
 : PRIOR FILING DATE: 2001-09-12
 : PRIOR APPLICATION NUMBER: 09/950,083
 : PRIOR FILING DATE: 2001-09-12
 : NUMBER OF SEQ ID NOS: 779
 : SOFTWARE: PatentIn ver. 2.0
 : SEQ ID NO 422
 : LENGTH: 9728
 : TYPE: DNA
 : ORGANISM: Homo sapiens
 PCT-US02-09105-422

Query Match	15.9%	Score	31.4	DB	1	length	9728
Best Local Similarity	48.1%	Pred	No. 18				
Matches	89	Conservative	0	Mismatches	96	Indels	0
						Gaps	0

QY	4	tegaagcaatggatttcgcgacttctctcagcttcacatcttcctctctgtgctcactc	63
Db	7085	ttctctctctctgggtacctggagctgagcagcagctgccaccatgcttggctgttttat	7026
QY	64	ttctctcttctctgtgatactctcaactctctgcgcgtgtccggtttccaaagacagtgcaa	123
Db	7025	tttttagttttttgtagagagctctacftatgttgccacagcctgctctcaaacctcttgaggctcc	6966
QY	124	tcatactactgcacaagaggaactcgttaagtgcacaaagatcgtgagctcgcgcgaagcgaaac	183
Db	6965	accacatcttccgcgcacagccttccctgagctgtaagctctggagattacaggtgccact	6906
QY	184	gtgtc	188
Db	6905	gtgcc	6901

Search completed: May 12, 2002, 21:35:57
Job time: 6853 sec

Search completed: May 12, 2002, 21:35:57
Job time: 6853 sec

/protein_id="CA870495.1"
/db_xref="GI:6779835"
/translation="MGFVLSQPSFLVSTLLFLVISHSCRAGSKKPVPIIYNRR
TGKQNM"
BASE COUNT 32 a 56 c 49 g 60 t
ORIGIN

Query Match 100.0%; Score 197; DB 6; Length 197;
Best Local Similarity 100.0%; Pred. No. 3.6e-49;
Matches 197; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ggcgcagcgcacgagtggttcgctcttccttcacgtccatcttccttcgtgtcta 60
DB 1 ggcgcagcgcacgagtggttcgctcttccttcacgtccatcttccttcgtgtcta 60
OY 61 cctctctcttccttcgtatctcctcactcttcgcgtgcggtcccaagaagccagtc 120
DB 61 cctctctcttccttcgtatctcctcactcttcgcgtgcggtcccaagaagccagtc 120
OY 121 caatcatcactgcaacagagagcagtcgaagtcgacagagagtgtagagtcgagagcg 180
DB 121 caatcatcactgcaacagagagcagtcgaagtcgacagagagtgtagagtcgagagcg 180
OY 181 aacgtgcagcagagtc 197
DB 181 aacgtgcagcagagtc 197

RESULT 2
LOCUS A98895 211 bp DNA linear PAT 26-JAN-2000
DEFINITION Sequence 3 from Patent WO9909189.
ACCESSION A98895
VERSION A98895.1 GI:6781856
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 211)
AUTHORS Derose, R. and Hoffmann, J.
TITLE GENE CODING FOR ANDROCTONINE, VECTOR CONTAINING SAME AND
JOURNAL TRANSFORMED DISEASE-RESISTANT PLANTS OBTAINED
DEROSE RICHARD (FR); HOFFMANN JULES (FR)
FEATURES
source Location/Qualifiers
1..211
CDS
/organism="unidentified"
/db_xref="taxon:32644"
/note="unnamed protein product"
/codon_start=1
/protein_id="CA870573.1"
/db_xref="GI:6781857"
/translation="MGFVLSQPSFLVSTLLFLVISHSCRARSVCKIKICRRRG
GCTYKCTNRPY"

BASE COUNT 33 a 58 c 57 g 63 t
ORIGIN

Query Match 59.6%; Score 117.4; DB 6; Length 211;
Best Local Similarity 77.0%; Pred. No. 4.9e-25;
Matches 161; Conservative 0; Mismatches 36; Indels 12; Gaps 1;

OY 1 ggcgcagcgcacgagtggttcgctcttccttcacgtccatcttccttcgtgtcta 60
DB 1 ggcgcagcgcacgagtggttcgctcttccttcacgtccatcttccttcgtgtcta 60
OY 61 cctctctcttccttcgtatctcctcactcttcgcgtgcggtcccaagaagccagtc 120
DB 61 cctctctcttccttcgtatctcctcactcttcgcgtgcggtcccaagaagccagtc 120
OY 61 cctctctcttccttcgtatctcctcactcttcgcgtgcggtcccaagaagccagtc 120
DB 61 cctctctcttccttcgtatctcctcactcttcgcgtgcggtcccaagaagccagtc 120
OY 121 caatcatcactgcaacagagagcagtcgaagtcgacagagagtgtagagtcgagagcg 168
DB 121 caatcatcactgcaacagagagcagtcgaagtcgacagagagtgtagagtcgagagcg 168

DB 121 tcacagatctgcacgagagagcggtgcttctactacaaagtcacagccactgag 180
OY 169 ctgcgcagcgcacgagtggttcgctcttccttcacgtccatcttccttcgtgtcta 197
DB 181 ctgcgcagcgcacgagtggttcgctcttccttcacgtccatcttccttcgtgtcta 209

RESULT 3
LOCUS A95926 106 bp DNA linear PAT 26-JAN-2000
DEFINITION Sequence 4 from Patent WO924594.
ACCESSION A95926
VERSION A95926.1 GI:6779832
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 106)
AUTHORS Derose, R. and Hoffmann, J.
TITLE GENE CODING FOR ANDROCTONINE, VECTOR CONTAINING SAME AND RESULTING
JOURNAL TRANSFORMED DISEASE-RESISTANT PLANTS
PATENT: WO 924594-A 4 20-MAY-1999;
DEROSE RICHARD (FR); HOFFMANN JULES (FR)
FEATURES
source Location/Qualifiers
1..106
CDS
/organism="unidentified"
/db_xref="taxon:32644"
/note="unnamed protein product"
/codon_start=1
/protein_id="CA870494.1"
/db_xref="GI:6779833"
/translation="MGFVLSQPSFLVSTLLFLVISHSCRA"

BASE COUNT 8 a 34 c 20 g 44 t
ORIGIN

Query Match 52.3%; Score 103; DB 6; Length 106;
Best Local Similarity 100.0%; Pred. No. 1.1e-20;
Matches 103; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ggcgcagcgcacgagtggttcgctcttccttcacgtccatcttccttcgtgtcta 60
DB 1 ggcgcagcgcacgagtggttcgctcttccttcacgtccatcttccttcgtgtcta 60
OY 61 cctctctcttccttcgtatctcctcactcttcgcgtgcggtcccaagaagccagtc 103
DB 61 cctctctcttccttcgtatctcctcactcttcgcgtgcggtcccaagaagccagtc 103

RESULT 4
LOCUS A98894 106 bp DNA linear PAT 26-JAN-2000
DEFINITION Sequence 2 from Patent WO9909189.
ACCESSION A98894
VERSION A98894.1 GI:6781854
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 106)
AUTHORS Derose, R. and Hoffmann, J.
TITLE GENE CODING FOR ANDROCTONINE, VECTOR CONTAINING SAME AND
JOURNAL TRANSFORMED DISEASE-RESISTANT PLANTS OBTAINED
DEROSE RICHARD (FR); HOFFMANN JULES (FR)
FEATURES
source Location/Qualifiers
1..106
CDS
/organism="unidentified"
/db_xref="taxon:32644"
/note="unnamed protein product"

/codon_start=1
/protein_id="CAB70572.1"
/db_xref="GI:6781855"
/translation="MGFVLFPSOLFVSTLLFLVISHSCRA"
BASE COUNT 8 a 34 c 20 g 44 t
ORIGIN

Query Match 52.3%; Score 103; DB 6; Length 106;
Best Local Similarity 100.0%; Pred. No. 1.1e-20;
Matches 103; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcgtgaacgacatgggttcgtgcttctctcagcttccatcttctctgctcga 60
|||||
DB 1 GCGTCAGACCCATGGGTTTCGTCTTCTCTCAGCTTCACATCTTCCCTTGTGTCTA 60
QY 61 cctctctctctctctgtgactctcactctgcgcgtgcgg 103
|||||
DB 61 CTCTCTCTCTTCTCTGTGATCTCTCACTCTGCCGCGCGG 103

RESULT 5
AX014749 120 bp DNA linear PAT 07-SEP-2000
LOCUS AX014749 Sequence 4 from Patent W09953053.
DEFINITION AX014749
ACCESSION AX014749.1 GI:10041017
VERSION
KEYWORDS
SOURCE
ORGANISM
synthetic construct.
artificial sequence.
1 (bases 1 to 120)
REFERENCE
AUTHORS
TITLE
JOURNAL
Lambert, M., Hofmann, J., Bulet, P. and Brookhart, G. L.
Gene coding for helminthocin and use thereof
Patent: WO 9953053-A 4 21-OCT-1999;
LAMBERTY MIRILE (FR); HOFMANN JULES (FR); BULET PHILIPPE (FR);
RHONE-POULENC AGROCHIMIE (FR); BROOKHART GARY LEE (US)

FEATURES
source
1..120
Location/Qualifiers
/organism="synthetic construct"
/db_xref="taxon:32630"
12..>101
/note="unnamed protein product"
/codon_start=1
/transl_table=11
/protein_id="CAC07580.1"
/db_xref="GI:10041018"
/translation="MGFVLFPSOLFVSTLLFLVISHSCRA"
BASE COUNT 14 a 36 c 23 g 47 t
ORIGIN

CDS

Query Match 51.6%; Score 101.6; DB 6; Length 120;
Best Local Similarity 92.2%; Pred. No. 2.9e-20;
Matches 107; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 gcgtgaacgacatgggttcgtgcttctctcagcttccatcttctctgctcga 60
|||||
DB 1 GCGTCAGACCCATGGGTTTCGTCTTCTCTCAGCTTCACATCTTCCCTTGTGTCTA 60
QY 61 cctctctctctctctgtgactctcactctgcgcgtgcgg 116
|||||
DB 61 CTCTCTCTCTTCTCTGTGATCTCTCACTCTGCCGCGCGG 116

RESULT 6
A95925 98 bp DNA linear PAT 26-JAN-2000
LOCUS A95925
DEFINITION Sequence 3 from Patent W09924594.
ACCESSION A95925
VERSION A95925.1 GI:6779830
KEYWORDS
SOURCE
unidentified.

ORGANISM unidentified
unclassified.
REFERENCE
1 (bases 1 to 98)
AUTHORS
TITLE
JOURNAL
DEROSE RICHARD (FR); HOFMANN JULES (FR)
Location/Qualifiers
1..98
/organism="unidentified"
/db_xref="taxon:32644"
<1..66
/note="unnamed protein product"
/codon_start=1
/protein_id="CAB70493.1"
/db_xref="GI:6779831"
/translation="GSKRPVPIIYCNRTKRCQM"
BASE COUNT 25 a 23 c 34 g 16 t
ORIGIN

Query Match 48.7%; Score 96; DB 6; Length 98;
Best Local Similarity 100.0%; Pred. No. 1.4e-18;
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 102 ggtccaaagaagccaggtccatctctcagcagagagactgtaagtccagagg 161
|||||
DB 1 GGTTCAGAGAGCCAGTCCATCTCTACTGCAACAGAGAGACTGTAGTCCAGAGG 60
QY 162 atgtgactcgcgcgagcgacgtgtcgcagcagatcc 197
|||||
DB 61 ATGTGAGCTCGCGCGAGCGAAGTGTGAGATCC 96

RESULT 7
AX014748 261 bp DNA linear PAT 07-SEP-2000
LOCUS AX014748 Sequence 3 from Patent W09953053.
DEFINITION AX014748
ACCESSION AX014748.1 GI:10041015
VERSION
KEYWORDS
SOURCE
ORGANISM
synthetic construct.
artificial sequence.
1 (bases 1 to 261)
REFERENCE
AUTHORS
TITLE
JOURNAL
Lambert, M., Hofmann, J., Bulet, P. and Brookhart, G. L.
Gene coding for helminthocin and use thereof
Patent: WO 9953053-A 3 21-OCT-1999;
LAMBERTY MIRILE (FR); HOFMANN JULES (FR); BULET PHILIPPE (FR);
RHONE-POULENC AGROCHIMIE (FR); BROOKHART GARY LEE (US)

FEATURES
source
1..261
Location/Qualifiers
/organism="synthetic construct"
/db_xref="taxon:32630"
3..227
/note="unnamed protein product"
/codon_start=1
/transl_table=11
/protein_id="CAC07579.1"
/db_xref="GI:10041016"
/translation="MGFVLFPSOLFVSTLLFLVISHSCRADKLIGSCWCAVNV
TSDNGCKRGRYKGGGSGFANVNCER"
BASE COUNT 39 a 66 c 74 g 82 t
ORIGIN

CDS

Query Match 47.4%; Score 93.4; DB 6; Length 261;
Best Local Similarity 98.9%; Pred. No. 9.5e-18;
Matches 94; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 10 ccatggttcgtgcttctctcagcttccatcttctgtgctactctctc 69
|||||

Db 1 CCATGGCTTTCCTTTCTCTCAGCTTCACATCTTTCCTTGTGCTACTCTCTTC 60
Qy 70 ttctcttgatctctcatctctgacgagc 104
|||||
Db 61 TTTTCCTGTGATCTCTCATCTCTGCGCGCAT 95

RESULT 8

LOCUS A95931 87 bp DNA linear PAT 26-JAN-2000
DEFINITION Sequence 9 from Patent WO9924594.
ACCESSION A95931
VERSION A95931.1 GI:6779839
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 87)
AUTHORS Derose,R. and Hoffmann,J.
TITLE GENE CODING FOR THANATIN, VECTOR CONTAINING SAME AND RESULTING
JOURNAL TRANSFORMED DISEASE-RESISTANT PLANTS
DEPOSE RICHARD (FR); HOFFMANN JULES (FR)
PATENT: WO 9924594-A 9 20-MAY-1999;
LOCATION/Qualifiers

FEATURES
source 1..87
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 14 a 31 c 21 g 21 t
ORIGIN

Query Match 43.1%; Score 85; DB 6; Length 87;
Best Local Similarity 100.0%; Pred. No. 3.1e-15;

Matches 85; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 113 gccagtcgaatcatctactgcacagagagactgtaagtgcagagagatgtgactcg 172
|||||
Db 87 GCCAGTGCATCATCTACTGTCACACAGAGACTGTGTAAGTCCAGAGATGTGAGCTCG 28
Qy 173 gccagggcaacgctgcagagatcc 197
|||||
Db 27 GCGAGCGCAACGTGTGACGCGATCC 3

RESULT 9

LOCUS A95928 75 bp DNA linear PAT 26-JAN-2000
DEFINITION Sequence 6 from Patent WO9924594.
ACCESSION A95928
VERSION A95928.1 GI:6779836
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 75)
AUTHORS Derose,R. and Hoffmann,J.
TITLE GENE CODING FOR THANATIN, VECTOR CONTAINING SAME AND RESULTING
JOURNAL TRANSFORMED DISEASE-RESISTANT PLANTS
DEPOSE RICHARD (FR); HOFFMANN JULES (FR)
PATENT: WO 9924594-A 6 20-MAY-1999;
LOCATION/Qualifiers

FEATURES
source 1..75
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 5 a 23 c 13 g 34 t
ORIGIN

Query Match 37.3%; Score 73.4; DB 6; Length 75;
Best Local Similarity 98.7%; Pred. No. 1e-11;
Matches 74; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 gcgcgacgacatgggttcgtctctctcctcagctccatcttctctctgtgtcta 60

Db 1 GCGTCGACGCGATGGTTCGTCCTTTCTCAGCTTCACATCTTTCCTTGTGCTCTA 60
Qy 61 ctctctctcttc 75
|||||
Db 61 CTCTCTCTCTTTC 75

RESULT 10

LOCUS A98896 75 bp DNA linear PAT 26-JAN-2000
DEFINITION Sequence 4 from Patent WO9909189.
ACCESSION A98896
VERSION A98896.1 GI:6781858
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 75)
AUTHORS Derose,R. and Hoffmann,J.
TITLE GENE CODING FOR ANDROCTONINE, VECTOR CONTAINING SAME AND
JOURNAL TRANSFORMED DISEASE-RESISTANT PLANTS OBTAINED
DEPOSE RICHARD (FR); HOFFMANN JULES (FR)
PATENT: WO 9909189-A 4 25-FEB-1999;
LOCATION/Qualifiers

FEATURES
source 1..75
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 5 a 23 c 13 g 34 t
ORIGIN

Query Match 37.3%; Score 73.4; DB 6; Length 75;
Best Local Similarity 98.7%; Pred. No. 1e-11;

Matches 74; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 gcgtcgacgcatgggttcgtctctctctcagctccatcttctctctgtgtcta 60
|||||
Db 1 GCGTCGACGCGATGGTTCGTCCTTTCTCAGCTTCACATCTTTCCTTGTGCTCTA 60
Qy 61 ctctctctcttc 75
|||||
Db 61 CTCTCTCTCTTTC 75

RESULT 11

LOCUS AX014750 75 bp DNA linear PAT 07-SEP-2000
DEFINITION Sequence 5 from Patent WO9953053.
ACCESSION AX014750
VERSION AX014750.1 GI:10041019
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct.
REFERENCE 1 (bases 1 to 75)
AUTHORS Lambery,M., Hofmann,J., Bulet,P. and Brookhart,G.L.
TITLE Gene coding for helioimicine and use thereof
JOURNAL LAMBERY MIREILLE (FR); HOFMANN JULES (FR); BULET PHILIPPE (FR);
RHONE POULENC AGRICHIMIE (FR); BROOKHART GARY LEE (US)
PATENT: WO 9953053-A 5 21-OCT-1999;
LOCATION/Qualifiers

FEATURES
source 1..75
/organism="synthetic construct"
/db_xref="taxon:32630"
BASE COUNT 5 a 23 c 13 g 34 t
ORIGIN

Query Match 37.3%; Score 73.4; DB 6; Length 75;
Best Local Similarity 98.7%; Pred. No. 1e-11;
Matches 74; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ggcgtcgaacgcatgggttcgtgcttctcctcagcttccatcttctctgtgtctta 60
|||||
Db 1 GCGTCGACCGCGAGGGCTTCGTCTTCCTCCAGCTCCATCTTCCTTCGTGCTCA 60
OY 61 cctctctctcttctc 75
|||||
Db 61 CTCTCTCTCTTCTTCC 75

RESULT 12

A95929/c
LOCUS A95929 72 bp DNA linear PAT 26-JAN-2000
DEFINITION Sequence 7 from Patent WO9924554.
ACCESSION A95829
VERSION A95929.1 GI:6779837
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS Derose,R. and Hoffmann,J.
TITLE GENE CODING FOR THANATIN, VECTOR CONTAINING SAME AND RESULTING TRANSFORMED DISEASE-RESISTANT PLANTS
JOURNAL Patent: WO 9924554-A 7 20-MAY-1999;
DEROSE RICHARD (FR); HOFFMANN JULES (FR)
FEATURES
source 1..72
/organism="unidentified"
/db_xref="taxon:32644"

BASE COUNT 33 a 11 c 23 g 5 t
ORIGIN

Query Match 34.2%; Score 67.4; DB 6; Length 72;
Best Local Similarity 98.6%; Pred. No. 6.6e-10;
Matches 68; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 35 gctccatcttctctctctgtgtctactcttctcttctcttctgtatctcactctg 94
|||||
Db 72 GCTTCGATCTTCCTTCCTGTGTGCTACTCTTCCTTCCTTCGTGATCTACTCTTG 13
OY 95 ccgtgcccgg 103
|||||
Db 12 CCGTGCCGG 4

RESULT 13

A98897/c
LOCUS A98897 72 bp DNA linear PAT 26-JAN-2000
DEFINITION Sequence 5 from Patent WO9909189.
ACCESSION A98897
VERSION A98897.1 GI:6781859
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS Derose,R. and Hoffmann,J.
TITLE GENE CODING FOR ANDROCTONINE, VECTOR CONTAINING SAME AND TRANSFORMED DISEASE-RESISTANT PLANTS OBTAINED
JOURNAL Patent: WO 9909189-A 5 25-FEB-1999;
DEROSE RICHARD (FR); HOFFMANN JULES (FR)
FEATURES
source 1..72
/organism="unidentified"
/db_xref="taxon:32644"

BASE COUNT 33 a 11 c 23 g 5 t
ORIGIN

Query Match 34.2%; Score 67.4; DB 6; Length 72;
Best Local Similarity 98.6%; Pred. No. 6.6e-10;
Matches 68; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 35 gctccatcttctctctgtgtctactcttcttcttcttctgtgacatctcactctg 94
|||||
Db 72 GCTTCGATCTTCCTTCCTGTGTGCTACTCTTCCTTCCTTCGTGATCTACTCTTG 13
OY 95 ccgtgcccgg 103
|||||
Db 12 CCGTGCCGG 4

RESULT 14

AX014751/c
LOCUS AX014751 72 bp DNA linear PAT 07-SEP-2000
DEFINITION Sequence 6 from Patent WO9953053.
ACCESSION AX014751
VERSION AX014751.1 GI:10041020
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS Lamberly,M., Hoffmann,J., Bulet,P. and Brookhart,G.L.
TITLE Gene coding for heliothine and use thereof
JOURNAL Patent: WO 9953053-A 6 21-OCT-1999;
LAMBERTY MIREILLE (FR); HOFFMANN JULES (FR); BULET PHILIPPE (FR); RHONE POULENC AGROCHIMIE (FR); BROOKHART GARY LEE (US)
FEATURES
source 1..72
/organism="synthetic construct"
/db_xref="taxon:32630"

BASE COUNT 33 a 11 c 23 g 5 t
ORIGIN
Query Match 34.2%; Score 67.4; DB 6; Length 72;
Best Local Similarity 98.6%; Pred. No. 6.6e-10;
Matches 68; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 35 gctccatcttctctctgtgtctactcttcttcttcttctgtgacatctcactctg 94
|||||
Db 72 GCTTCGATCTTCCTTCCTGTGTGCTACTCTTCCTTCCTTCGTGATCTACTCTTG 13
OY 95 ccgtgcccgg 103
|||||
Db 12 CCGTGCCGG 4

RESULT 15

A95924
LOCUS A95924 63 bp DNA linear PAT 26-JAN-2000
DEFINITION Sequence 2 from Patent WO9924554.
ACCESSION A95924
VERSION A95924.1 GI:6779828
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS Derose,R. and Hoffmann,J.
TITLE GENE CODING FOR THANATIN, VECTOR CONTAINING SAME AND RESULTING TRANSFORMED DISEASE-RESISTANT PLANTS
JOURNAL Patent: WO 9924554-A 2 20-MAY-1999;
DEROSE RICHARD (FR); HOFFMANN JULES (FR)
FEATURES
source 1..63
/organism="unidentified"
/db_xref="taxon:32644"

CDS

/note="unnamed protein product"
/codon_start=1
/protein_id="CAB70492.1"
/db_xref="GI:6779828"
/translation="GSKRPVPIYCNRRGKQRM"

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 12, 2002, 18:58:33 ; Search time 333.08 Seconds
(without alignments) 170.104 Million cell updates/sec

Title: US-09-554-024-1

Perfect score: 33
Sequence: 1 atcattactgcacagagagactgtaagtc 33

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_032802:*

- 1: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1980.DAT:*
- 2: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1981.DAT:*
- 3: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1982.DAT:*
- 4: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1983.DAT:*
- 5: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1984.DAT:*
- 6: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1985.DAT:*
- 7: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1986.DAT:*
- 8: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1987.DAT:*
- 9: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1988.DAT:*
- 10: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1989.DAT:*
- 11: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1990.DAT:*
- 12: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1991.DAT:*
- 13: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1992.DAT:*
- 14: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1993.DAT:*
- 15: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1994.DAT:*
- 16: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1995.DAT:*
- 17: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1996.DAT:*
- 18: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:*
- 19: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1998.DAT:*
- 20: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1999.DAT:*
- 21: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:*
- 22: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:*
- 23: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:*
- 24: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	33	100.0	33	20	AAx59770
2	33	100.0	63	20	AAx59771
3	33	100.0	98	20	AAx59772
4	33	100.0	197	20	AAx59774
5	22.4	67.9	3102	17	AAx15600
6	20.2	61.2	589	22	AAH87754
7	20.2	61.2	18073	20	AAV83948
8	20	60.6	1275	23	AAx82520
9	20	60.6	1737	23	AAx82525

10	20	60.6	2289	23	AAx89143	DNA encoding novel
11	19.8	60.0	179	23	AAx48172	Enterococcus faeca
12	19.6	59.4	404	21	AAx26094	Human secreted pro
13	19.4	58.8	792	23	ABL21645	Drosophila melanog
14	19.4	58.8	1488	22	AAx83638	Novel human transp
15	19.4	58.8	1626	22	AAx83639	Novel human transp
16	19.4	58.8	1977	22	AAx83635	Novel human transp
17	19.4	58.8	2115	22	AAx83636	Novel human transp
18	19.4	58.8	2448	22	AAx83644	Novel human transp
19	19.4	58.8	2586	22	AAx83645	Novel human transp
20	19.4	58.8	2700	22	AAx83640	Novel human transp
21	19.4	58.8	2937	22	AAx83641	Novel human transp
22	19.4	58.8	3075	22	AAx83642	Novel human transp
23	19.4	58.8	3189	23	AAx83637	Novel human transp
24	19.4	58.8	3219	23	ABL21644	Drosophila melanog
25	19.4	58.8	3660	22	AAx83646	Novel human transp
26	19.4	58.8	4149	22	AAx83643	Novel human transp
27	19.4	58.8	4341	22	AAx49454	D. melanogaster la
28	19.4	58.8	5020	22	AAx83647	Novel human transp
29	19.4	58.8	5718	22	AAx57077	CDNA encoding Dros
30	19.4	58.8	5718	23	ABL04853	Drosophila melanog
31	19.4	58.8	9400	23	ABL04852	Drosophila melanog
32	19.2	58.2	380	21	AAx6207	Pinus radiata tran
33	19.2	58.2	393	21	AAx57329	Pinus radiata tran
34	19.2	58.2	5053	21	AAx51296	A. terreus ORF1 en
35	19.2	58.2	16595	22	AAx33411	DNA encoding human
36	19.2	58.2	33000	21	AAx51294	A. terreus D48 seg
37	19.2	58.2	532	22	AAx72049	Human immune/Haem
38	19.2	57.6	258	23	AAx53542	Haemophilus influe
39	18.8	57.0	485	22	AAx06825	Human brain expres
40	18.8	57.0	584	15	AAx073858	Borrelia 2591 Ospe
41	18.8	57.0	824	16	AAx51447	Outer surface prot
42	18.8	57.0	1352	22	AAx39721	Genomic sequence #
43	18.8	57.0	3249	22	ABx15833	Human nervous syst
44	18.8	57.0	3249	22	AAx03071	Human reproductive
45	18.8	57.0	3249	22	AAx03071	Human reproductive

ALIGNMENTS

RESULT 1	
AAx59770	standard; cDNA, 33 BP.
XX	
XX	AAx59770;
XX	
XX	26-JUL-1999 (first entry)
XX	
XX	CDNA encoding a thanatine peptide.
DE	
XX	Thanatine; fungal disease; bacterial disease; Cercospora bettula;
KW	Cladosporium herbarum; Fusarium curvum; F. graminearum;
KW	Phytophthora cinamomi; selection marker; plant transformation;
KW	herbicide resistance; ss.
XX	
OS	Psodius sp.
XX	
PN	FR2770853-A1.
XX	
PD	14-MAY-1999.
XX	
PF	07-NOV-1997; 97FR-0014263.
XX	
PR	07-NOV-1997; 97FR-0014263.
XX	
PA	(RHON) RHONE-POULENC AGROCHIMIE.
XX	
PI	Derose R, Freyssinet G, Hoffmann J;
XX	WPI, 1999-315645/27.
DR	P-PsDB; AAT13462.
XX	

PF New nucleic acid encoding thanatine useful as a selection marker for
PT transformation of plants

XX Claim 3; Page 13; 24pp; French.

CC The specification describes a nucleic acid sequence containing the
CC sequence that encodes thanatine. Plants transformed with this nucleic
CC acid sequence are resistant to fungal and bacterial diseases,
CC specifically those caused by *Cercospora beticola*, *Cladosporium*
CC *herbarum*, *Fusarium cumorum*, *F. graminearum* and *Phytophthora*
CC *cinamonii*. The nucleic acid sequence may also be used as a selection
CC marker for transformation of plants with other coding sequences,
CC e.g. those that impart resistance to herbicides. The present sequence
CC encodes the peptide thanatine.

SQ Sequence 33 BP; 10 A; 7 C; 9 G; 7 T; 0 other;

Query Match 100.0%; Score 33; DB 20; Length 33;

Best Local Similarity 100.0%; Pred. No. 0.00014; Mismatches 0; Indels 0; Gaps 0;

Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 atcattactgcacagagagactgtaagtgc 33
DB 1 atcattactgcacagagagagactgtaagtgc 33

RESULT 2

AA59771 ID AAX59771 standard; cDNA; 63 BP.

XX AAX59771;

DT 26-JUL-1999 (first entry)

XX cDNA encoding a thanatine peptide.

DE Thanatine: fungal disease; bacterial disease; *Cercospora beticola*;

KM *Cladosporium herbarum*; *Fusarium cumorum*; *F. graminearum*;

KW *Phytophthora cinamonii*; selection marker; plant transformation;

KW herbicide resistance; ss.

OS *Psodius* sp.

XX FR2770853-A1.

XX 14-MAY-1999.

PF 07-NOV-1997; 97FR-0014263.

XX 07-NOV-1997; 97FR-0014263.

PA (RHON) RHONE-POULENC AGROCHIMIE.

PI Derose R, Freyssinet G, Hoffmann J;

DR WPI: 1999-315645/27.

XX P-PSDB; AAY15463.

PT New nucleic acid encoding thanatine useful as a selection marker for

PT transformation of plants

PS Claim 4; Page 13; 24pp; French.

XX The specification describes a nucleic acid sequence containing the
CC sequence that encodes thanatine. Plants transformed with this nucleic
CC acid sequence are resistant to fungal and bacterial diseases,
CC specifically those caused by *Cercospora beticola*, *Cladosporium*
CC *herbarum*, *Fusarium cumorum*, *F. graminearum* and *Phytophthora*
CC *cinamonii*. The nucleic acid sequence may also be used as a selection
CC marker for transformation of plants with other coding sequences,
CC e.g. those that impart resistance to herbicides. The present sequence
CC encodes the peptide thanatine.

XX Sequence 63 BP; 19 A; 14 C; 19 G; 11 T; 0 other;

Query Match 100.0%; Score 33; DB 20; Length 63;

Best Local Similarity 100.0%; Pred. No. 0.00016; Mismatches 0; Indels 0; Gaps 0;

Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 atcattactgcacagagagactgtaagtgc 33
DB 22 atcattactgcacagagagagactgtaagtgc 54

RESULT 3

AA59772 ID AAX59772 standard; cDNA; 98 BP.

XX AAX59772;

DT 26-JUL-1999 (first entry)

XX cDNA encoding the signal peptide of the PR-1a gene of tobacco.

DE Thanatine: fungal disease; bacterial disease; *Cercospora beticola*;

KM *Cladosporium herbarum*; *Fusarium cumorum*; *F. graminearum*;

KW *Phytophthora cinamonii*; selection marker; plant transformation;

KW herbicide resistance; PR-1a gene; tobacco; ss.

OS *Nicotiana tabacum*.

XX FR2770853-A1.

XX 14-MAY-1999.

PF 07-NOV-1997; 97FR-0014263.

XX 07-NOV-1997; 97FR-0014263.

PA (RHON) RHONE-POULENC AGROCHIMIE.

PI Derose R, Freyssinet G, Hoffmann J;

DR WPI: 1999-315645/27.

XX P-PSDB; AAY15464.

PT New nucleic acid encoding thanatine useful as a selection marker for

PT transformation of plants

PS Example 1; Page 14; 24pp; French.

XX The specification describes a nucleic acid sequence containing the
CC sequence that encodes thanatine. Plants transformed with this nucleic
CC acid sequence are resistant to fungal and bacterial diseases,
CC specifically those caused by *Cercospora beticola*, *Cladosporium*
CC *herbarum*, *Fusarium cumorum*, *F. graminearum* and *Phytophthora*
CC *cinamonii*. The nucleic acid sequence may also be used as a selection
CC marker for transformation of plants with other coding sequences,
CC e.g. those that impart resistance to herbicides. The present sequence
CC encodes the signal peptide of the PR-1a gene of tobacco, which was
CC used to make fusion products of the invention.

SQ Sequence 98 BP; 25 A; 23 C; 34 G; 16 T; 0 other;

Query Match 100.0%; Score 33; DB 20; Length 98;

Best Local Similarity 100.0%; Pred. No. 0.00017; Mismatches 0; Indels 0; Gaps 0;

Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 atcattactgcacagagagactgtaagtgc 33
DB 22 atcattactgcacagagagagactgtaagtgc 54

RESULT 4
AAK59774 standard; cDNA; 197 BP.
XX
AC AAK59774;
XX
DT 26-JUL-1999 (first entry)
XX
DE cDNA encoding a thanatine fusion product of the invention.
XX
KW Thanatine; fungal disease; bacterial disease; Cercospora beticola;
KW Cladosporium herbarum; Fusarium culmorum; F. graminearum;
KW Phytophthora cinamomi; selection marker; plant transformation;
KW herbicide resistance; PR-1a gene; tobacco; ss.
XX
OS Synthetic.
XX
FN FR2770853-A1.
XX
PD 14-MAY-1999.
XX
PE 07-NOV-1997; 97FR-0014263.
XX
PR 07-NOV-1997; 97FR-0014263.
XX
PA (RHON) RHONE-POULENC AGROCHIMIE.
XX
PI Derose R, Freyssinet G, Hoffmann J;
XX
XX WPI: 1999-315645/27.
DR P-PSDB; AAY15466.
XX
PT New nucleic acid encoding thanatine useful as a selection marker for
XX transformation of plants
XX
PS Claim 8; Page 14-15; 24pp; French.
XX
XX The specification describes a nucleic acid sequence containing the
CC sequence that encodes thanatine. Plants transformed with this nucleic
CC acid sequence are resistant to fungal and bacterial diseases,
CC specifically those caused by Cercospora beticola, Cladosporium
CC herbarum, Fusarium culmorum, F. graminearum and Phytophthora
CC cinamomi. The nucleic acid sequence may also be used as a selection
CC marker for transformation of plants with other coding sequences,
CC e.g. those that impart resistance to herbicides. The present sequence
CC encodes a thanatine fusion product of the invention, comprising the
CC signal peptide of the PR-1a gene of tobacco.
XX
SQ Sequence 197 BP; 32 A; 56 C; 49 G; 60 T; 0 other;

Query Match 100.0%; Score 33; DB 20; Length 197;
Best Local Similarity 100.0%; Pred. No. 0.00019;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atcacctactgcaacagagagactgtaagtc 33
|||
Db 123 atcacctactgcaacagagagactgtaagtc 155

RESULT 5
AAK15600
ID AAT15600 standard; DNA; 3102 BP.
XX
AC AAT15600;
XX
DT 07-APR-1996 (first entry)
XX
DE Laccase-LCC3 gene.
XX
KW Laccase-LCC3; Polyporus pinsitus; Trametes villosa; primer;
KW polymerase chain reaction; PCR; signal peptide; cellulase;
KW Aspergillus oryzae; cDNA probe; Escherichia coli; plasmid pDSY20;

KW Lignin; lignosulphonate; polymerisation; Kraft pulp;
KW depolymerisation; oxidation; hair dye; phenol; aniline; vector;
KW cloning; Basidiomycetes; ds.
XX
XX Polyporus pinsitus.
OS Trametes villosa.
XX
FH Key
FH CAAAT_signal
FT 14..18
FT Location/Qualifiers
FT /*tag= a
FT 406..410
FT TATA_signal
FT /*tag= b
FT 483..2695
FT CDS
FT /*tag= c
FT /product= Laccase-LCC3
FT /note= "EC-1.10.3.2"
FT 483..545
FT sig_peptide
FT /*tag= d
FT 483..665
FT exon
FT /*tag= e
FT 666..720
FT Intron
FT /*tag= f
FT 721..789
FT exon
FT /*tag= g
FT 790..845
FT Intron
FT /*tag= h
FT 846..965
FT exon
FT /*tag= i
FT 966..1020
FT Intron
FT /*tag= j
FT 1021..1134
FT exon
FT /*tag= k
FT 1135..1192
FT Intron
FT /*tag= l
FT 1193..1258
FT exon
FT /*tag= m
FT 1259..1318
FT Intron
FT /*tag= n
FT 1319..1399
FT exon
FT /*tag= o
FT 1400..1460
FT Intron
FT /*tag= p
FT 1461..1616
FT exon
FT /*tag= q
FT 1617..1672
FT Intron
FT /*tag= r
FT 1673..1873
FT exon
FT /*tag= s
FT 1874..1929
FT Intron
FT /*tag= t
FT 1930..1986
FT exon
FT /*tag= u
FT 1987..2036
FT Intron
FT /*tag= v
FT 2037..2237
FT exon
FT /*tag= w
FT 2238..2296
FT Intron
FT /*tag= x
FT 2297..4213
FT exon
FT /*tag= y
FT 2414..2469
FT Intron
FT /*tag= z
FT 2470..2586
FT exon
FT /*tag= {
FT 2587..2638
FT Intron
FT /*tag= |
FT 2639..2695
FT exon
FT /*tag= }
XX
XX
PN WO9600290-A1.
XX
PD 04-JAN-1996.
XX

PF 15-JUN-1995; 95MO-US07536.
 XX
 PR 15-MAY-1995; 95US-0441147.
 PR 24-JUN-1994; 94US-0255534.
 XX
 PA (NOVO) NOVO NORDISK BIOTECH INC.
 XX (NOVO) NOVO-NORDISK AS.
 PI Asslyng DA, Dalboge H, Schneider P, Xu F, Yaver DS;
 XX WPI; 1996-068874/07.
 DR P-PSDB; AAR90723.
 XX
 PT DNA constructs for expression of Polyporus laccase enzymes - for use
 PT in e.g. lignin manipulation, juice mfr., phenol polymerisation and
 PT phenol resin prodn
 XX
 PS Claim 7; Page 67-70; 137pp; English.
 XX
 CC The sequence encodes laccase-LCC3 (PI 5.23) from Polyporus pinsutus
 CC (Trametes villosa). Polymerase chain reaction (PCR) amplification of
 CC P. pinsutus cDNA using primers 3331 (AA115603) and 3332 (AA115604) gives
 CC a 1500 bp fragment, which is joined to a signal peptide sequence
 CC from a 43-kDa cellulase using primer PHD433 (AA115605) and a pUC
 CC forward primer in PCR. Clones are expressed in Aspergillus oryzae,
 CC and a cDNA probe is obtained and used to screen a P. pinsutus
 CC genomic library in Escherichia coli DH5-alpha, giving plasmid
 CC pDSX20 (24GEN), with an 8.5-kb HindIII insert (NRRL-B-21267).
 CC Screening also results in isolation of LCC1 (AA115598), LCC2
 CC (AA115599), LCC4 (AA115601) and LCC5 (AA115602), which encode different
 CC laccases produced by P. pinsutus. The laccases may be used to
 CC polymerise lignin or lignosulphonates, to depolymerise Kraft pulp,
 CC to oxidise dyes or precursors, in hair dye compositions, or to
 CC polymerise or oxidise a phenolic or aniline compound. These new
 CC laccases are well-expressed in Aspergillus spp. (with vector
 CC integration in the genome), in contrast to previous basidiomycete
 CC laccases, which give low yields of recombinant enzyme.
 XX
 SQ Sequence 3102 BP; 669 A; 935 C; 758 G; 739 T; 1 other;
 Query Match 67.9%; Score 22.4; DB 17; Length 3102;
 Best Local Similarity 81.2%; Pred. No. 9.1;
 Matches 26; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
 Oy . 2 tcactactgcacacagggagactgtaagtc 33
 Db 2623 tcactcttcgacacagggagactgtaacaaagc 2654
 RESULT 6
 AAH87754/C
 ID AAH87754 standard; cDNA; 589 BP.
 XX
 AC AAH87754;
 XX
 XX 25-SEP-2001 (first entry)
 DE Peppermint plant oil gland expressed cDNA 110.
 XX
 KW Peppermint; plant oil gland cell; terpenoid essential oil; resin;
 KW genetic mapping; antisense suppression; recombinant expression; ss.
 XX
 OS Mentha x piperita.
 XX
 PN WO200153319-A1.
 XX
 XX 26-JUL-2001.
 PD
 XX 19-JAN-2001; 2001WO-US02567.
 PF
 XX 20-JAN-2000; 2000US-0177264.
 PR
 XX

PA (CROT/) CROTEAU R B.
 PA (LANG/) LANGE B M.
 PA (WILD/) WILDUNG M R.
 XX
 PI Croteau RB, Lange BM, Wildung MR;
 XX WPI; 2001-488706/53.
 DR
 XX
 PT New nucleic acid molecules corresponding to mRNA molecules expressed in
 PT peppermint oil glands for enhancing expression of plant oil gland cell
 PT proteins -
 XX
 PS Claim 1; Page 119-120; 251pp; English.
 XX
 CC The invention relates to nucleic acid molecules (AAH87645-AAH88116) that
 CC correspond to all or part of a mRNA molecule expressed in plant oil
 CC gland cells, especially peppermint and plant oil glands that produce
 CC terpenoid essential oils and resins. The nucleic acids are useful for
 CC genetically mapping a plant genome for genes expressed in plant oil
 CC gland cells and to suppress (for example by antisense suppression) or
 CC enhance their expression (for example by genetically transforming a
 CC plant cell with a replicable expression vector that expresses one or more
 CC proteins naturally expressed in plant oil gland cells). The nucleic acids
 CC are also useful for recombinant expression of plant oil gland proteins
 CC required for terpenoid essential oil and/or resin production in bacterial
 CC and/or yeast cells.
 XX
 SQ Sequence 589 BP; 164 A; 115 C; 135 G; 175 T; 0 other;
 Query Match 61.2%; Score 20.2; DB 22; Length 589;
 Best Local Similarity 75.8%; Pred. No. 56;
 Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
 Oy . 1 atcatctacgcacacagggagactgtaagtc 33
 Db 396 ATCATCTCACAGGAAACACACACACACGTAACACG 364
 RESULT 7
 AAV83948/C
 ID AAV83948 standard; DNA; 18073 BP.
 XX
 AC AAV83948;
 XX
 XX 03-MAR-1999 (first entry)
 DE Bacterial artificial chromosome (BAC)-F2 contig 8.
 XX
 KW Yeast artificial chromosome; YAC; probe; eukaryotic chromosome;
 KW neocentromere; replication; extra-chromosomal element; segregation;
 KW cell division; artificial chromosome; gene therapy; BAC; transgenic;
 KW human artificial chromosome; bacterial artificial chromosome; ss.
 XX
 OS Synthetic.
 XX
 PN WO9851790-A1.
 XX
 XX 19-NOV-1998.
 PD
 XX 13-MAY-1998; 98MO-AU00352.
 PF
 XX 26-AUG-1997; 97AU-0008791.
 PR 13-MAY-1997; 97AU-0006784.
 XX
 PA (AMRA-) AMRAD OPERATIONS PTY LTD.
 XX
 XX Cancellia MR, Choo K, Du Sart D;
 PI
 XX WPI; 1999-009773/01.
 DR
 XX
 PT 'New isolated nucleic acid comprising neocentromere sequences from
 PT eukaryotic chromosome - used to produce replicable, segregating

PT artificial chromosomes that can carry large amounts of DNA for gene
PT therapy
XX
PS Claim 10; Page 209-222; 540pp; English.
XX
CC The present sequence represents a bacterial artificial chromosome (BAC)
CC contig, and exemplifies the invention. The specification describes
CC nucleic acid sequences derived from a eukaryotic chromosome, including a
CC neocentromere or its functional derivative or hybrid, that are able, in
CC a compatible cell, of replicating, acting as extra-chromosomal element
CC and segregating during cell division. The sequences can be used to
CC construct artificial chromosomes for use in gene therapy comprising a
CC replicable, segregating nucleic acid that confers a specific phenotype
CC on cells. Human artificial chromosomes can propagate in human cells and
CC carry large amounts of DNA (e.g. therapeutic genes), and, being
CC extra-chromosomal, they are not mutagenic. The artificial chromosomes
CC are also useful for generation of transgenic plants and animals, in
CC production of proteins and to make diagnostic reagents, e.g. for
CC expression of cytokines, receptors and growth factors, or to increase
CC the copy number of a gene in a cell. The constructs may also be
CC used for functional and structural analysis of chromosomes.
XX
SQ Sequence 18073 BP; 4755 A; 3584 C; 3589 G; 6145 T; 0 other:

Query Match 61.2%; Score 20.2; DB 20; Length 18073;
Best Local Similarity 88.0%; Pred. No. 1e+02; Mismatches 0; Gaps 0;
Matches 22; Conservative 0; Indels 3; Indels 0; Gaps 0;

OY 8 actgcacacagagagactgtaagt 32
||| ||||| ||||| ||||| ||
Db 8535 ACTGTTACAGAGAGACTGTAATG 8511

RESULT 8
AAS82520
ID AAS82520 standard; cDNA; 1275 BP.
XX
AC AAS82520;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #18324.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI: 2001-639362/73.
DR P-PSDB: ABG18333.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1; SEQ ID No 18324; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and

CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1275 BP; 293 A; 328 C; 361 G; 293 T; 0 other:

Query Match 60.6%; Score 20; DB 23; Length 1275;
Best Local Similarity 82.1%; Pred. No. 79;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 4 attactgcacacagagagactgtaagt 31
||||| ||||| ||| || ||||| ||
Db 205 attacttcaactgcagagagcgttaagt 232

RESULT 9
AAS82525
ID AAS82525 standard; cDNA; 1737 BP.
XX
AC AAS82525;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #18329.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI: 2001-639362/73.
DR P-PSDB: ABG18338.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1; SEQ ID No 18329; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,

CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcr_sequences.

SO Sequence 1737 BP; 449 A; 405 C; 462 G; 421 T; 0 other;

Query Match 60.6%; Score 20; DB 23; Length 1737;
Best Local Similarity 82.1%; Pred. No. 83;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 4 atctactgcacagagagactgtaagt 31
||||| ||||| ||| ||| |||||
Db 988 atctactcactcagagagcgttaagt 1015

RESULT 10

AAS89143 standard; CDNA; 2289 BP.

XX AAS89143:

DT 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #24947.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;

KM food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX Homo sapiens.

XX WO200175067-A2.

XX 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

DR WPI; 2001-639362/73.

DR P-PSDB; ABG24956.

XX New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations

PT responsible for genetic disorders or other traits and to assess

PT biodiversity

XX Claim 1; SEQ ID No 24947; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and

CC polypeptide (II) sequences. (I) is useful as hybridisation probes,

CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome

CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcr_sequences.

SO Sequence 2289 BP; 516 A; 595 C; 678 G; 500 T; 0 other;

Query Match 60.6%; Score 20; DB 23; Length 2289;
Best Local Similarity 82.1%; Pred. No. 87;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 4 atctactgcacagagagactgtaagt 31
||||| ||||| ||| ||| |||||
Db 493 atctactcactcagagagcgttaagt 520

RESULT 11

AAS48172 standard; DNA; 179 BP.

XX AAS48172:

DT 13-FEB-2002 (first entry)

DE Enterococcus faecalis cellular proliferation inhibitory sequence #742.

XX Antisense; ss; prokaryotic cellular proliferation;

KM antibiotic; antibacterial; drug design.

XX Enterococcus faecalis.

XX WO200170955-A2.

XX 27-SEP-2001.

PF 21-MAR-2001; 2001WO-US09180.

PR 21-MAR-2000; 2000US-191078P.

PR 23-MAY-2000; 2000US-206848P.

PR 26-MAY-2000; 2000US-207727P.

PR 23-OCT-2000; 2000US-242578P.

PR 27-NOV-2000; 2000US-253625P.

PR 22-DEC-2000; 2000US-257931P.

PR 16-FEB-2001; 2001US-269308P.

XX (ELIT-) ELITRA PHARM INC.

PI Haselbeck R, Ohlsen KU, Zyskind JW, Wall D, Trawick JD, Carr GJ;

PI Yamamoto RT, Xu HH;

DR WPI; 2001-611495/70.

XX New polynucleotides for the identification and development of

PT antibiotics, comprise sequences of antisense nucleic acids -

XX Claim 1; Seq ID No 749; 511pp; English.

XX The invention relates to antisense inhibitors of genes essential to

CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella*
CC *pneumoniae*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence is an antisense
CC oligonucleotide of the invention.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pctl_sequences.
XX
XX
SQ Sequence 179 BP; 54 A; 38 C; 30 G; 57 T; 0 other;

Query Match 60.0%; Score 19.8; DB 23; Length 179;
Best Local Similarity 77.4%; Pred. No. 67;
Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 1 atcattactgcacagcagcagcgttaagt 31
|||||
DB 43 atcattactgcacagcagcagcgtatacgt 73

RESULT 12
AAC26094/C
ID AAC26094 standard; CDNA: 404 BP.
XX
XX AAC26094;

DT 06-OCT-2000 (first entry)

DE Human secreted protein 5' EST, SEQ ID NO: 30169.

XX Human: 5' EST; expressed sequence tag; secreted protein; CDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX

OS Homo sapiens.

XX EPI033401-A2.

XX 06-SEP-2000.

XX 21-FEB-2000; 2000EP-0200610.

XX 26-FEB-1999; 99US-0122487.

XX (GEST) GENSET.

XX Dumas Milne Edwards J, Duclert A, Giordano J;

XX WPI: 2000-500381/45.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
XX obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
XX diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX Claim 1; SEQ ID 30169; 71pp + CD-ROM; English.

XX The present sequence is one of a large number of 5' ESTs derived from
XX mRNAs encoding secreted proteins. No ORF has yet been conclusively
XX identified within the present sequence. The 5' ESTs were prepared from
XX total human RNAs or polyA⁺ RNAs derived from 30 different tissues. EST
XX sequences usually correspond mainly to the 3' untranslated region (UTR)
XX of the mRNA because they are often obtained from oligo-dT primed cDNA
XX libraries. Such ESTs are not well suited for isolating cDNA sequences

CC derived from the 5' ends of mRNAs and even in those cases where longer
CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC They are used to obtain upstream regulatory sequences and to design
CC expression and secretion vectors.
XX
XX
SQ Sequence 404 BP; 132 A; 62 C; 72 G; 136 T; 2 other;

Query Match 59.4%; Score 19.6; DB 21; Length 404;
Best Local Similarity 84.6%; Pred. No. 94;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 3 cactactgcacagcagcagcgtgta 28
|||||
DB 39 CATCTGCTGCACAGCAGCATGCTGTA 14

RESULT 13
ABL21645/C
ID ABL21645 standard; DNA: 792 BP.
XX
XX ABL21645;

DT 26-MAR-2002 (first entry)

DE *Drosophila melanogaster* genomic polynucleotide SEQ ID NO 16408.

XX *Drosophila*; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX

OS *Drosophila melanogaster*.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI: 2001-656860/75.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from *Drosophila* and for elucidating cell signalling and cell-cell
XX interactions -
XX Claim 1; SEQ ID NO 16408; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from *Drosophila*. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins
XX (ABB57737-ABB72072).
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pctl_sequences.

XX Sequence 792 BP; 200 A; 258 C; 190 G; 144 T; 0 other;

Query Match 58.8%; Score 19.4; DB 23; Length 792;
Best Local Similarity 79.3%; Pred. No. 1.3e+02;

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 12, 2002, 18:13:38 ; Search time 1854.92 Seconds
(without alignments)

372,294 Million cell updates/sec

Title: US-09-554-024-1

Perfect score: 1 atcattactgcaacagagagactgttaagtgc 33

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

GenEmbl: *
1: gb_ba: *
2: gb_htg: *
3: gb_in: *
4: gb_on: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_sts: *
12: gb_sy: *
13: gb_un: *
14: gb_vi: *
15: em_ba: *
16: em_fun: *
17: em_hum: *
18: em_in: *
19: em_mu: *
20: em_on: *
21: em_or: *
22: em_ov: *
23: em_pat: *
24: em_pl: *
25: em_ro: *
26: em_sts: *
27: em_un: *
28: em_vi: *
29: em_htg_hum: *
30: em_htg_inv: *
31: em_htg_other: *
32: em_htgo_inv: *
33: em_htgo_inv: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
------------	-------	-------------	--------	----	-------------

1	100.0	33	6	A95923	Sequence 1
2	100.0	63	6	A95924	Sequence 2
3	100.0	87	6	A95931	Sequence 9
4	100.0	98	6	A95925	Sequence 3
5	100.0	197	6	A95927	Sequence 5
6	67.9	540	14	AF170051	HIV-2 Iso
7	67.9	3102	6	165231	Sequence 5
8	67.9	157853	2	AC095159	Rattus no
9	64.8	5138	2	AF090432	Danio rer
10	64.8	5355	5	AF229450	Danio rer
11	64.8	5469	5	AF229449	Danio rer
12	64.8	81357	2	AC005792	Homo sapi
13	64.8	93553	2	AC010593	Homo sapi
14	64.8	155221	9	AC016587	Homo sapi
15	63.6	234	6	163568	Sequence 14
16	63.6	88645	2	AC105715	Rattus no
17	63.6	100061	2	AC016551	Homo sapi
18	63.6	106554	2	RN258C08	Rattus no
19	63.6	159732	2	RN40B13	Rattus no
20	63.6	161092	2	AC100725	Mus muscu
21	63.6	170569	2	AC108109	Homo sapi
22	63.6	18721	2	RN462P8	Rattus no
23	63.6	191132	9	AC010282	Homo sapi
24	63.6	268200	1	AP000061	Aeropyrum
25	63.0	1341	9	HSRSDB11	Homo sapi
26	63.0	20803	3	AF045644	Caenorhab
27	63.0	32986	3	CE12K783	Caenorhabd
28	63.0	37255	3	U40933	Caenorhabd
29	63.0	69479	8	AC008597	Homo sapi
30	63.0	91720	8	F2J10	Sequence
31	63.0	156277	9	AC026795	Homo sapi
32	63.0	162949	2	AC068029	Homo sapi
33	63.0	176621	9	AC024564	Homo sapi
34	63.0	192645	2	AC006753	Caenorhab
35	62.4	129285	1	AC091726	Gallus ga
36	61.8	14936	2	AB076402	Halomonas
37	61.8	105613	2	AC097777	Rattus no
38	61.8	210115	9	AL442127	Human DNA
39	61.2	18073	6	AX033920	Sequence
40	61.2	42313	3	U13645	Caenorhabd
41	61.2	99816	2	AL139144	Homo sapi
42	61.2	126679	2	HS0534B8	Human DNA
43	61.2	150465	8	AC091749	Oryza sat
44	61.2	175345	2	AC022301	Mus muscu
45	61.2	180746	2	AL138803	Homo sapi

ALIGNMENTS

RESULT	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	FEATURES	CDS
1	A95923	Sequence 1 from Patent WO924594.	A95923	A95923.1	GI:6779826	unidentified, unclassified.	1 (bases 1 to 33)	Derose, R. and Hoffmann, J.	TRANSFORMED DISEASE-RESISTANT PLANTS	Patent: WO 924594-A 1 20-MAY-1999;	DEROSE RICHARD (FR); HOFFMANN JULES (FR)	Location/Qualifiers

1..33
/organism="unidentified"
/gb_xref="taxon:32644"
<1..>33
/note="unnamed protein product"
/codon_start=1

BASE COUNT 10 a 7 c 9 g 7 t
ORIGIN /protein_id="CAB70491.1"
/db_xref="GI:6779827"
/translation="IYCNRRFGKC"

Query Match 100.0%; Score 33; DB 6; Length 33;
Best Local Similarity 100.0%; Pred. No. 0.00013;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 atcactctactgcacagaggagactgtaagtgc 33
|||||
Db 1 ATCATCTACTGCACAGAGGACTGTAAGTGC 33

RESULT 2
LOCUS A95924 63 bp DNA linear PAT 26-JAN-2000
DEFINITION Sequence 2 from Patent WO924594.
ACCESSION A95924
VERSION A95924.1 GI:6779828
KEYWORDS
SOURCE
ORGANISM
unidentified.
unclassified.

REFERENCE 1 (bases 1 to 63)
AUTHORS Derose, R. and Hoffmann, J.
TITLE GENE CODING FOR THANATIN, VECTOR CONTAINING SAME AND RESULTING
TRANSFORMED DISEASE-RESISTANT PLANTS
JOURNAL Patent: WO 924594-A 2 20-MAY-1999;
DEROSE RICHARD (FR); HOFFMANN JULES (FR)
Location/Qualifiers

FEATURES
source 1..63
/organism="unidentified"
/db_xref="taxon:32644"

CDS
/note="unnamed protein product"
/codon_start=1
/protein_id="CAB70492.1"
/db_xref="GI:6779829"
/translation="GSKRPVPIIYCNRRFGKCQRM"

BASE COUNT 19 a 14 c 19 g 11 t
ORIGIN

Query Match 100.0%; Score 33; DB 6; Length 63;
Best Local Similarity 100.0%; Pred. No. 0.00014;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 atcactctactgcacagaggagactgtaagtgc 33
|||||
Db 22 ATCATCTACTGCACAGAGGACTGTAAGTGC 54

RESULT 3
LOCUS A95931 87 bp DNA linear PAT 26-JAN-2000
DEFINITION Sequence 9 from Patent WO924594.
ACCESSION A95931
VERSION A95931.1 GI:6779839
KEYWORDS
SOURCE
ORGANISM
unidentified.
unclassified.

REFERENCE 1 (bases 1 to 87)
AUTHORS Derose, R. and Hoffmann, J.
TITLE GENE CODING FOR THANATIN, VECTOR CONTAINING SAME AND RESULTING
TRANSFORMED DISEASE-RESISTANT PLANTS
JOURNAL Patent: WO 924594-A 9 20-MAY-1999;
DEROSE RICHARD (FR); HOFFMANN JULES (FR)
Location/Qualifiers

FEATURES
source 1..87
/organism="unidentified"

BASE COUNT 14 a 31 c 21 g 21 t
ORIGIN /organism="unidentified"
/db_xref="taxon:32644"

Query Match 100.0%; Score 33; DB 6; Length 87;
Best Local Similarity 100.0%; Pred. No. 0.00015;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 atcactctactgcacagaggagactgtaagtgc 33
|||||
Db 77 ATCATCTACTGCACAGAGGACTGTAAGTGC 45

RESULT 4
LOCUS A95925 98 bp DNA linear PAT 26-JAN-2000
DEFINITION Sequence 3 from Patent WO924594.
ACCESSION A95925
VERSION A95925.1 GI:6779830
KEYWORDS
SOURCE
ORGANISM
unidentified.
unclassified.

REFERENCE 1 (bases 1 to 98)
AUTHORS Derose, R. and Hoffmann, J.
TITLE GENE CODING FOR THANATIN, VECTOR CONTAINING SAME AND RESULTING
TRANSFORMED DISEASE-RESISTANT PLANTS
JOURNAL Patent: WO 924594-A 3 20-MAY-1999;
DEROSE RICHARD (FR); HOFFMANN JULES (FR)
Location/Qualifiers

FEATURES
source 1..98
/organism="unidentified"
/db_xref="taxon:32644"

CDS
/note="unnamed protein product"
/codon_start=1
/protein_id="CAB70493.1"
/db_xref="GI:6779831"
/translation="GSKRPVPIIYCNRRFGKCQRM"

BASE COUNT 25 a 23 c 34 g 16 t
ORIGIN

Query Match 100.0%; Score 33; DB 6; Length 98;
Best Local Similarity 100.0%; Pred. No. 0.00015;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 atcactctactgcacagaggagactgtaagtgc 33
|||||
Db 22 ATCATCTACTGCACAGAGGACTGTAAGTGC 54

RESULT 5
LOCUS A95927 197 bp DNA linear PAT 26-JAN-2000
DEFINITION Sequence 5 from Patent WO924594.
ACCESSION A95927
VERSION A95927.1 GI:6779834
KEYWORDS
SOURCE
ORGANISM
unidentified.
unclassified.

REFERENCE 1 (bases 1 to 197)
AUTHORS Derose, R. and Hoffmann, J.
TITLE GENE CODING FOR THANATIN, VECTOR CONTAINING SAME AND RESULTING
TRANSFORMED DISEASE-RESISTANT PLANTS
JOURNAL Patent: WO 924594-A 5 20-MAY-1999;
DEROSE RICHARD (FR); HOFFMANN JULES (FR)
Location/Qualifiers

FEATURES
source 1..197
/organism="unidentified"

CDS
/db_xref="taxon:32644"
12.167
/note="unnamed protein product"
/codon_start=1
/protein_id="CA870495.1"
/db_xref="GI:6779835"
/translation="MGFVLPQLPSFLVSTLLFLVISHSCRAAGKPPVPIYCNRR
TGCCQRN"

BASE COUNT 32 a 56 c 49 g 60 t
ORIGIN

Query Match 100.0%; Score 33; DB 6; Length 197;
Best Local Similarity 100.0%; Pred. No. 0.00016;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 atcactctacacagagagactgtaagtc 33
Db 123 ATCATCTACTGCACAGAGAGACTGTAAGTC 155

RESULT 6
AF170051 540 bp DNA linear VRL 05-MAR-2001
LOCUS HIV-2 isolate 96207 from France envelope glycoprotein (env) gene,
DEFINITION partial cds.
ACCESSION AF170051
VERSION AF170051.1 GI:11990555
KEYWORDS Human immunodeficiency virus type 2.
SOURCE Human immunodeficiency virus type 2.
ORGANISM Viruses; Retroviridae; Lentivirus; Primate
leutivirus group.
1 (bases 1 to 540)
REFERENCE
AUTHORS Diamond, P., Aprel, C., Robertson, D. L., Souquiere, S., Lepretre, A.,
Tateno, Y., Plantier, J. C., Brun-Vezinet, F., and Simon, F.
TITLE Variability of human immunodeficiency virus type 2 (HIV-2)
infecting patients living in France
JOURNAL Virology 280 (1), 19-30 (2001)
PUBMED 11162815
AUTHORS 2 (bases 1 to 540)
Francis, S., Diamond, P., Aprel, C., Robertson, D. L., Souquiere, S.,
Lepretre, A., Tateno, Y., and Brun-Vezinet, F.
TITLE Direct Submissions
Submitted (19-JUL-1999) Laboratoire de Virologie et Services des
Maladies Infectieuses et Tropicales, Hôpital Bichat - Claude
Bernard, Paris 75877 Paris Cedex 18, France
FEATURES
source location/Qualifiers
1.540
/organism="Human immunodeficiency virus type 2"
/isolate="96207"
/db_xref="taxon:11709"
/country="France"
/gene="env"
/gene_start=1
/gene_end=540
/product="envelope glycoprotein"
/protein_id="AAG42191.1"
/db_xref="GI:11990556"
/translation="DWSTFETSIKPCVKILPLCVAMNCTKPSNTTATASNTSTT
STATCAVAVNETPCVSTNCTGEEEMITCNSTGLODKKRRNREYTSQDYC
ESNNIDGNISRCYMSHCNTVTSTSCDKHMDTRRCAPGVIALRCNDYNTG
FAVNCSSVVAATCTRMETQ"

BASE COUNT 190 a 115 c 110 g 125 t
ORIGIN

Query Match 67.9%; Score 22.4; DB 14; Length 540;
Best Local Similarity 81.2%; Pred. No. 16;
Matches 26; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 atcactctacacagagagactgtaagtc 32
Db 132 ATCAACACTGCACACTGCAGAGCTGTAATG 163

RESULT 7
165231 3102 bp DNA linear PAT 07-OCT-1997
LOCUS Sequence 5 from patent US 5667531.
DEFINITION I65231.1 GI:2481801
ACCESSION I65231
VERSION I65231.1 GI:2481801
KEYWORDS Unknown.
SOURCE Unknown.
ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 3102)
AUTHORS Yaver, D. Sue., Xu, F., Dalb.o slashed.ge, H., Schneider, P. and
Aaslyng, D. A.
TITLE Dye compositions containing purified polyporus laccases and nucleic
acids encoding same
JOURNAL Patent: US 5667531-A 5 16-SEP-1997;
FEATURES Location/Qualifiers
1.3102
/organism="unknown"
BASE COUNT 669 a 935 c 758 g 739 t 1 others
ORIGIN

Query Match 67.9%; Score 22.4; DB 6; Length 3102;
Best Local Similarity 81.2%; Pred. No. 19;
Matches 26; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 2 tcatctactgcaacagagagactgtaagtc 33
Db 2623 TCATCTTTTGCACAGAGAGACTGCAACAAGC 2654

RESULT 8
AC095159 157853 bp DNA linear HTG 20-DEC-2001
LOCUS Rattus norvegicus clone CH230-8C21, *** SEQUENCING IN PROGRESS ***
DEFINITION 62 unordered pieces.
ACCESSION AC095159
VERSION AC095159.2 GI:17942168
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 157853)
REFERENCE
AUTHORS Muzny, D. M., Adams, C., Adio-Oduola, B., Ali-oshman, F. R., Allen, C.,
Alsbrooks, S. L., Amaral, H. C., Are, J. R., Banks, T., Barbara, J.,
Benton, J., Bimaga, K., Blankenburg, K., Bonin, D., Bouch, J.,
Bowie, S., Briley, M., Brown, E., Brown, M., Bryant, N. P., Buhay, C.,
Burch, P., Burkett, C., Burrell, K. L., Byrd, N. C., Carron, T. F.,
Carter, M., Cavazos, S. R., Chacko, J., Chavez, D., Chen, G., Chen, R.,
Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C. D., Cox, C.,
Coyle, M. D., Dathorne, S. R., David, R., Davila, M. L., Davis, C.,
Day-Carroll, L., Dederich, D. A., Delaney, K. R., Delgado, O.,
Denn, A. L., Ding, Y., Dinh, H. H., Douthwaite, K. J., Draper, H.,
Duan, R., Rocha, S., Durbin, K. J., Eamhart, C., Edgar, D., Edwards, C. C.,
Elhaj, C., Escotto, M., Falls, T., Ferriguto, D., Flagg, N., Ford, J.,
Foster, P., Franz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T.,
Garza, N., Gill, R., Gorrell, J. H., Guevara, W., Gunaratne, P., Hale, S.,
Hamilton, K., Harris, C., Harris, K., Hart, M., Haylak, P., Hawes, A.,
Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Hollway, C.,
Hollins, B., Homsi, F., Howard, S., Huber, J., Huylk, S., Hume, J.,
Jackson, L. E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S.,
Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J.,
Kovar, C., Kravovic, J., Kuresh, A., Landry, N., Leal, B., Lewis, L. C.,
Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, C., Liu, W.,
Loulseged, H., Lozano, R. J., Lu, X., Lucier, A., Lucier, R., Luna, R.,

Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A.,
Martinez, E., Massey, E., Mawlin, E., McLeod, M.P., Meador, M.,
Mel, G., Metzger, M., Miner, G., Miner, Z., Mitchell, T., Mohabhat, K.,
Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N.,
Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenko, S.,
Oguni, M., Okunolu, G., Oragunye, N., Oviator, R., Pace, A., Payton, B.,
Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Fu, L.,
Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Rolfe, M.,
Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shooshtari, N.,
Stinson, I., Sodergren, E., Sonalike, T., Sparks, A., Stanley, H.,
Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K.,
Tang, H., Tansey, J., Taylor, C., Taylor, P., Telford, B., Thomas, R.,
Thomas, S., Umani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R.,
Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C.,
Washington, S., Williams, G., Williamson, A., Wlarczyk, R., Woodson, S.,
Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
Weinstock, G. and Gibbs, R.

Direct Submission
2 (bases 1 to 157853)
Unpublished
Worley, K.C.
Submitted (16-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Dec 20, 2001 this sequence version replaced g1:15625713.

Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu

Project Information
Center project name: GCvH
Center clone name: CH230-8C21

Summary Statistics
Assembly program: Phrap; version 0.990329first call to
findhaplist
Consensus quality: 124710 bases at least Q40
Consensus quality: 132941 bases at least Q30
Consensus quality: 139599 bases at least Q20
Estimated insert size: 121782; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-gel estimation
Quality coverage: 2.1x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_data_format.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 62 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1
10810 10809: contig of 10809 bp in length
*
* 10910 10909: gap of unknown length
*
* 16008 16007: contig of 5098 bp in length
*
* 16108 16107: gap of unknown length
*
* 20378 20377: contig of 4270 bp in length
*
* 20478 20477: gap of unknown length
*
* 25693 25692: contig of 5215 bp in length
*
* 25793 25792: gap of unknown length
*
* 31441 31441: contig of 5549 bp in length
*
* 31442 31441: gap of unknown length
*
* 38548 38548: contig of 7107 bp in length
*
* 38549 38548: gap of unknown length
*
* 42504 42503: contig of 3855 bp in length
*
* 42604 42603: gap of unknown length
*
* 46321 46321: contig of 3618 bp in length
*
* 46322 46321: gap of unknown length
*
* 50780 50780: contig of 4459 bp in length
*
* 50880 50880: gap of unknown length
*
* 54670 54670: contig of 3790 bp in length
*
* 54671 54670: gap of unknown length
*
* 54771 54770: contig of 3106 bp in length
*
* 57877 57876: gap of unknown length
*
* 57977 57976: contig of 3404 bp in length
*
* 61381 61380: gap of unknown length
*
* 61481 61480: gap of unknown length
*
* 63946 63945: contig of 2465 bp in length
*
* 63946 63945: gap of unknown length
*
* 64046 64045: contig of 1826 bp in length
*
* 65872 65871: gap of unknown length
*
* 65972 65971: contig of 3336 bp in length
*
* 69368 69367: gap of unknown length
*
* 69468 69467: contig of 3105 bp in length
*
* 72573 72572: gap of unknown length
*
* 72673 72672: contig of 3121 bp in length
*
* 75794 75793: gap of unknown length
*
* 75894 75893: contig of 1989 bp in length
*
* 77883 77882: gap of unknown length
*
* 77983 77982: contig of 1734 bp in length
*
* 79717 79716: gap of unknown length
*
* 79817 79816: gap of unknown length
*
* 82887 82886: contig of 3071 bp in length
*
* 82888 82887: gap of unknown length
*
* 82988 82987: contig of 2225 bp in length
*
* 85213 85212: gap of unknown length
*
* 85313 85312: contig of 1757 bp in length
*
* 87070 87069: gap of unknown length
*
* 87170 87169: gap of unknown length
*
* 89967 89966: contig of 2797 bp in length
*
* 90067 90066: gap of unknown length
*
* 92128 92128: contig of 2062 bp in length
*
* 92129 92128: gap of unknown length
*
* 92229 92228: gap of unknown length
*
* 94503 94503: contig of 2275 bp in length
*
* 94603 94603: gap of unknown length
*
* 97620 97620: contig of 3017 bp in length
*
* 97721 97720: gap of unknown length
*
* 97721 97720: contig of 2718 bp in length
*
* 100439 100438: gap of unknown length
*
* 100539 100538: gap of unknown length
*
* 102723 102722: contig of 2184 bp in length
*
* 102823 102822: gap of unknown length
*
* 104843 104843: contig of 2021 bp in length
*
* 104844 104843: gap of unknown length
*
* 107122 107122: contig of 2179 bp in length
*
* 107123 107122: gap of unknown length
*
* 109547 109547: contig of 2355 bp in length
*
* 109548 109547: gap of unknown length
*
* 109648 109647: contig of 1508 bp in length
*
* 111156 111155: gap of unknown length
*
* 111256 111255: gap of unknown length
*
* 113579 113579: contig of 2324 bp in length
*
* 113679 113679: gap of unknown length
*
* 115794 115794: contig of 2115 bp in length
*
* 115894 115894: gap of unknown length
*
* 115895 115894: contig of 2078 bp in length
*
* 117972 117972: gap of unknown length
*
* 118072 118072: contig of 1672 bp in length
*
* 119744 119744: gap of unknown length
*
* 119745 119744: gap of unknown length
*
* 119845 119844: gap of unknown length
*
* 121303 121302: contig of 1458 bp in length
*
* 121403 121402: gap of unknown length
*
* 122766 122766: contig of 1364 bp in length
*
* 122866 122866: gap of unknown length
*
* 124357 124357: contig of 1491 bp in length
*
* 124358 124357: gap of unknown length
*
* 124457 124457: gap of unknown length
*
* 125917 125917: contig of 1460 bp in length
*
* 125918 125917: gap of unknown length
*
* 126018 126017: gap of unknown length
*
* 127778 127778: contig of 1761 bp in length
*
* 127878 127878: gap of unknown length
*
* 129566 129566: contig of 1688 bp in length
*
* 129666 129666: gap of unknown length
*
* 131623 131623: contig of 1957 bp in length
*
* 131624 131623: gap of unknown length
*
* 131724 131723: gap of unknown length
*
* 132811 132811: contig of 1188 bp in length
*
* 133011 133011: gap of unknown length
*
* 134523 134523: contig of 1512 bp in length
*
* 134623 134623: gap of unknown length
*
* 136223 136223: contig of 1600 bp in length
*
* 136224 136223: gap of unknown length
*
* 136323 136323: gap of unknown length

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

*	136324	137533:	contig of 1210 bp in length
*	137534	137633:	gap of unknown length
*	137634	138659:	contig of 1026 bp in length
*	138659	138759:	gap of unknown length
*	138760	139989:	contig of 1230 bp in length
*	139989	140089:	gap of unknown length
*	140090	141463:	contig of 1374 bp in length
*	141464	141563:	gap of unknown length
*	141564	143038:	contig of 1475 bp in length
*	143039	143138:	gap of unknown length
*	143139	144159:	contig of 1021 bp in length
*	144160	144259:	gap of unknown length
*	144260	145363:	contig of 1104 bp in length
*	145364	145463:	gap of unknown length
*	145464	146750:	contig of 1287 bp in length
*	146751	146850:	gap of unknown length
*	146851	147999:	contig of 1149 bp in length
*	148000	148099:	gap of unknown length
*	148100	149878:	contig of 1779 bp in length
*	149879	149978:	gap of unknown length

Query Match
Best Local Similarity 83.3%; Score 22; DB 2; Length 157853;
Matches 25; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 atcattactgcaacagagagctgtaag 30
Db 27901 ACCATATGACGAAACAGAGACTGGAAG 27930

RESULT 9
LOCUS AF090432/c 5138 bp mRNA linear VRT 28-DEC-1998
DEFINITION Danio rerio serrateb mRNA, complete cds.
ACCESSION AF090432
VERSION AF090432.1 GI:4063638
KEYWORDS zebrafish.
SOURCE Danio rerio
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.

REFERENCE
AUTHORS Haddon, C., Jiang, Y.J., Smathers, L. and Lewis, J.
TITLE 1 (bases 1 to 5138)
Jagged Genes Isolation, Characterization and Expression Analysis of Zebrafish
Delta-Notch signaling and the patterning of sensory cell
differentiation in the zebrafish ear: evidence from the mind bomb
mutant

JOURNAL Development 125 (23), 4637-4644 (1998)
MEDLINE 99026054
AUTHORS Jiang, Y.-J.
TITLE 2 (bases 1 to 5138)
JOURNAL Direct Submission
Submitted (04-SEP-1998) Vertebrate Development Laboratory, Imperial
Cancer Research Fund, 44 Lincoln's Inn Fields, London WC2A 3PX, UK

FEATURES
source location/Qualifiers

1..5138
/organism="Danio rerio"
/db_xref="taxon:7955"
/clone_lib="B. Appel"
/dev_stage="15-19 hour"
131..3895
/codon_start=1
/product="serrateb"
/protein_id="AAC8354.1"
/db_xref="GI:4063639"
/translation="MNMCIIRIRNMLPIACLLITMWTIKVSQSSGYFEQLIAYENVNG
LMDGCCDSTRNSODRCVDECDYFKVCLKEYSVTTTGCTFGSGSPDLVGGNI
ISKTKNSPKSTSDGKIIIPHFAMPBSYTLLEAMDMNSNONGEENLIERHNI
ASVWPGDHMSQIRHGTITAHIEIRIVRCDEYVYSGKNKQCRPRDYGHRCDPS
GNIVCLDMMGEBCRAICOGCNLHGCAVPECKCNMGQCFDECIPYGCGLH
GTCVMPDQCTCKRNWGLCDKDLNITGTHPCVNGGTCNMSPEDEINCAPEG1SGK
NCEIAPHACVSPNCANGCTCHEVPTGECNCPMVEGPTCAKMDDEACASPCAGGTC
IDLENGFEVCPQWVGKTCQIDLANECMGRPCVNAHSCKNMIGYHCDCEFGMAGQNC

DINLNGCHGOQNGCATCKELVHGGRHCCPAGFYGLHCEVRNKKCASGPQNGRCHV ILDSFVCEBPSNVNMLCEVESLSHPNCEPBNPCONTALCYSLPDEFCAPDEVEK TCERRKHCKMTPCQVYIDSCITIAVANSNSDGVRIHNSVGCPRDCLISQREGNCT CELEFTYTCHEENVNDCVSNPCRNNGTIDISSFOCPDWBEDLCSINNESSRS PKNGKGVLDVNFYCECANGMKGKTGHSRESDSSTCSNGGTCYHGDAFRCACP PGMGSTNCNTAKNSTGACPCNLNGCTGSHSRESDSSTCSNGGTCYHGDAFRCACP CYNGICVDGVNMFRECECAPGAGPDCRINIDECOSPCAYGATVDEINGRVCPL GRGPOCOEFYIGTICTHYAGLOPHGSRMEECNCCGVNGKCECTRYVGRKPCDL PGTRGRELHYCPRGRECELEHNFITCLSPCHOMGCSRPENPTQTCERPSYLDK SCARITLIFNDKLPYGTIVNVCSELNMPATRSALADHKLVLICDLSYKQDVEV AISOHDEPSSHQIOEASTIVSLSRHNSVWALVIEVKVEPVSQPDVILPV LCVVLCLVWFECVIVCWMTTRKRREKREKREKREKREKREKREKREKREKREK DAOERIKLMSPTPCNSGDDEDEDEDEDELEVEEMGTEGSKHPVOKYSRGAARTK NGLICTRSTSGSSPSLKAAYMGTFSPKDNCKVNNATAGOEKHCY"
--

BASE COUNT 1296 a 1224 c 1360 g 1258 t
Query Match
Best Local Similarity 64.8%; Score 21.4; DB 5; Length 5138;
Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 2 tcatctactgcaacagagagctgtaagtg 32
Db 4908 TTATGATCAGCAACAGAGACTGTAAGG 4878

RESULT 10
LOCUS AF229450/c 5355 bp mRNA linear VRT 28-SEP-2001
DEFINITION Danio rerio jagged2 (jag2) mRNA, complete cds, alternatively
spliced.
ACCESSION AF229450
VERSION AF229450.1 GI:15799277
KEYWORDS zebrafish.
SOURCE Danio rerio
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.

REFERENCE
AUTHORS Oda, T. and Chandrasekharappa, S.C.
TITLE 1 (bases 1 to 5355)
Isolation, Characterization and Expression Analysis of Zebrafish
Jagged Genes
Unpublished
2 (bases 1 to 5355)
JOURNAL Direct Submission
Submitted (31-JAN-2000) GMBB/NHGRI, NIH, 49 Convent Drive,
Bethesda, MD 20892-4442, USA

JOURNAL
MEDLINE
AUTHORS
TITLE
JOURNAL
Submitted (31-JAN-2000) GMBB/NHGRI, NIH, 49 Convent Drive,
Bethesda, MD 20892-4442, USA

FEATURES
source location/Qualifiers
1..5355
/organism="Danio rerio"
/db_xref="taxon:7955"
/chromosome="LGXX"
1..5355
/gene="jag2"
148..3798
/gene="jag2"
/note="Notch ligand; alternatively spliced"
/codon_start=1
/product="jagged2"
/protein_id="AAL08215.1"
/db_xref="GI:15799278"

/translation="MNMCIIRIRNMLPIACLLITMWTIKVSQSSGYFEQLIAYENVNG
LMDGCCDSTRNSODRCVDECDYFKVCLKEYSVTTTGCTFGSGSPDLVGGNI
ISKTKNSPKSTSDGKIIIPHFAMPBSYTLLEAMDMNSNONGEENLIERHNI
ASVWPGDHMSQIRHGTITAHIEIRIVRCDEYVYSGKNKQCRPRDYGHRCDPS
GNIVCLDMMGEBCRAICOGCNLHGCAVPECKCNMGQCFDECIPYGCGLH
GTCVMPDQCTCKRNWGLCDKDLNITGTHPCVNGGTCNMSPEDEINCAPEG1SGK
NCEIAPHACVSPNCANGCTCHEVPTGECNCPMVEGPTCAKMDDEACASPCAGGTC
IDLENGFEVCPQWVGKTCQIDLANECMGRPCVNAHSCKNMIGYHCDCEFGMAGQNC
EVSRRKCASGPQNGRCHVILDSFVCEBPSNVNMLCEVESLSHPNCEPBNPCONTA

Db	4816	TTATGATCAGCAGACAGAGACTGCTAAGCG	4786
Query Match	64.8%	Score 21.4	DB 5; Length 535;
Best Local Similarity	80.6%	Pred. No. 58;	
Matches	25; Conservative	0; Mismatches	6; Indels 0; Gaps 0;
2	tcattctactgacagagagagactgtaagt 32		
1			
Db	4816	TTATGATCAGCAGACAGAGACTGCTAAGCG	4786
RESULT 11	AF229449	5469 bp	mRNA linear VRT 28-SEP-2001
LOCUS	AF229449/c		
DEFINITION	Danio rerio jagged2 (jag2) mRNA, complete cds, alternatively		
LOCATION	spliced		
ACCESSION	AF229449		
VERSION	AF229449.1	GI:15799275	
KEYWORDS			
SOURCE	zebrafish.		
ORGANISM	Danio rerio		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.		
AUTHORS	Oda, T. and Chandrasekharappa, S.C.		
TITLE	Isolation, Characterization and Expression Analysis of Zebrafish Jagged Genes		
JOURNAL	Unpublished		
REFERENCE	2, (bases 1 to 5469)		
AUTHORS	Oda, T. and Chandrasekharappa, S.C.		
TITLE	Direct Submission		
JOURNAL	Submitted (31-JAN-2000) GMB/NHGRI, NIH, 49 Convent Drive, Bethesda, MD 20892-4442, USA		
FEATURES	Location/Qualifiers		
Source	1..5469		
gene	/organism="Danio rerio"		
CDs	/db_xref="taxon:7955"		
	/chromosome="LGX"		
	1..5469		
	/gene="jag2"		
	148..3912		
	/gene="jag2"		
	/note="Notch ligand; alternatively spliced"		
	/codon_start=1		
	/product="jagged2"		
	/protein_id="AAU08214.1"		
	/db_xref="GI:15799276"		
	/translation="MNMCIIRIRNLPIACLLITMMTRYSSSGYFELQILAVENNGECS LMQECDSPTRNODCRVRCDECTYRCKLKEVQSEYTGCTGCTPGSGDVLGNGIN FSRTAKNSPKSTSDVCKIITPFHPANRSTLTLEBAMDNDSTONGNEEULIRHIIH ASVNPGDHMOSIRHPGITAHIERIVRCDENYGSKCNQCRPDYFHYKCDPS GNIVCLDGMWDSIDRPIATICKQCNLIHGCGAVPECKNMGQCFDECIAPYGCILH GTCAVMPQCECKNMWGLICDKDLNLTGTHHPCVNGGCMNSPEDEYACAPESYSG NCEIAFHACVSNPCANGGTGHEVPTGFEGHCPCPMEGPTCKAMDDECASSPCAGGNC DILNENGECVCPQWVGKTCQIDANDEMGKPCVAHSGKNNIGSYHDCFGMAGGON DILNENGHCQGNKATCKELVHGYSKCCPAGFGLHCEVSRNKCASCPQNGRCRH ILDSFVECPSTNAGMLCEVESHHPNCEBPNQNTALCTSLPGDFCACPEDEEG		

Query Match	Best Local Similarity	64.8%	Score 21.4	DB 5	Length 5469
Matches 25	Conservative	0	Mismatches	6	Indels
0	0	0	Gaps	0	0
Oy	2	tcattctactgaacagagagactgtaagt	32		
Db	4930	TTATGATCAGCAACGAGAGCTGTAGG	4900		
RESULT 12					
AC005792					
LOCUS					
DEFINITION					
AC005792					
ACCESSION					
VERSION					
KEYWORDS					
SOURCE					
ORGANISM					
REFERENCE					
AUTHORS					
COMMENT					
JOURNAL					
REFERENCE					
AUTHORS					
JOURNAL					
COMMENT					
FEATURES					
Source					

```
repeat_region /clone.lib="Cal tech CIT-B BAC library"
765..1075
/rpt_family="AluJo"
complement(1824..1846)
repeat_region /rpt_family="AT-rich"
complement(1912..1997)
/rpt_family="7SLRNA"
complement(2009..2309)
/rpt_family="AluY"
complement(2312..2531)
/rpt_family="7SLRNA"
complement(3052..3144)
/misc_feature /note="predicted exon, program: grail2exons_human_1.3,
frame: 0, quality: good, score: 50.000"
repeat_region /rpt_family="AluJo"
complement(4576..4873)
/rpt_family="AluJo"
complement(4888..5416)
/rpt_family="LINE2"
complement(5428..5654)
/rpt_family="AluJo"
5655..5956
/rpt_family="AluY"
complement(5959..6039)
/rpt_family="AluJo"
complement(6571..6888)
/rpt_family="AluSx"
complement(7532..7568)
/rpt_family="(TAAAn)"
complement(7569..7872)
/rpt_family="AluSx"
complement(8178..8220)
/rpt_family="(CA)n"
complement(8221..8510)
/rpt_family="Aluub"
8741..8945
/note="predicted exon, program: grail2exons_human_1.3,
frame: 2, quality: good, score: 52.000"
repeat_region /rpt_family="MIR"
complement(11130..11264)
/rpt_family="FLAM_C"
12397..13128
/rpt_family="L1PB1"
complement(13140..13162)
/rpt_family="AT-rich"
complement(13323..13617)
/rpt_family="L1MB4"
complement(13670..13715)
/rpt_family="MIR"
complement(14717..15020)
/rpt_family="AluJo"
complement(15207..15253)
/rpt_family="Alu"
15264..15352
/rpt_family="LINE2"
complement(15352..15649)
/rpt_family="AluY"
complement(15740..16136)
/rpt_family="MSRA"
complement(16332..16629)
/rpt_family="AluJo"
complement(16934..17299)
/rpt_family="MLT1D"
complement(17315..17577)
/rpt_family="AluY"
complement(17589..17886)
/rpt_family="AluSx"
complement(17896..17974)
/rpt_family="MLT1D"
complement(18949..19125)
/misc_feature /note="BLASTN similarity to Z22632 (87..2633); match: 1,
score: 4.0e-65; database searched: nt; H.sapiens repeat
region DNA."
```

```
repeat_region complement(19132..19428)
/rpt_family="AluSx"
complement(20228..20268)
/rpt_family="(TAAAn)"
complement(20269..20570)
/rpt_family="AluSx"
20898..21065
/rpt_family="MSTB"
21066..21295
/rpt_family="AluSx"
21313..21479
/rpt_family="MSTB"
complement(21480..21785)
/rpt_family="AluJo"
21787..21851
/rpt_family="MSTB"
complement(23323..23618)
/rpt_family="AluSx"
24242..24346
/misc_feature /note="predicted exon, program: grail2exons_human_1.3,
frame: 0, quality: good, score: 73.000"
repeat_region /rpt_family="AT-rich"
complement(24579..24599)
/rpt_family="AT-rich"
complement(24877..24915)
/rpt_family="AT-rich"
complement(25821..25945)
/rpt_family="L1MB1"
25951..26234
/rpt_family="AluJo"
complement(26235..26372)
/rpt_family="L1MB3"
complement(26746..26798)
/rpt_family="MLT2B"
26803..26843
/rpt_family="(TA)n"
complement(26845..27250)
/rpt_family="MLT2B"
27332..27668
/rpt_family="THE1C"
complement(27671..27972)
/rpt_family="AluSx"
27976..28014
/rpt_family="THE1C"
29127..29427
/rpt_family="Aluub"
29444..29716
/rpt_family="AluJo"
complement(29995..30356)
/rpt_family="THE1B"
30392..30536
/rpt_family="MIR"
complement(30630..30933)
/rpt_family="AluJo"
30967..31333
/rpt_family="THE1B"
31385..31480
/rpt_family="MIR"
31575..31870
/rpt_family="AluSx"
complement(32080..32358)
/rpt_family="MLT1B"
complement(32365..32669)
/rpt_family="AluJo"
complement(32672..32841)
/rpt_family="MLT1B"
complement(33501..33799)
/rpt_family="AluSx"
33866..33991
/misc_feature /note="predicted exon, program: grail2exons_human_1.3,
frame: 1, quality: good, score: 50.000"
repeat_region /rpt_family="MIR"
complement(34438..34738)
repeat_region
```

repeat_region /rpt_family="AluX"
complement(34739..34767)
repeat_region /rpt_family="AT-rich"
complement(34779..34886)
repeat_region /rpt_family="FRAM"
34976..35221
repeat_region /rpt_family="LINE2"
35223..35528
repeat_region /rpt_family="AluX"
35529..35769
repeat_region /rpt_family="LINE2"
complement(36178..36478)
repeat_region /rpt_family="AluX"

Query Match 64.8%; Score 21.4; DB 9; Length 81357;
Best Local Similarity 80.6%; Pred. No. 77;
Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 2 tcatctactgcacagagagactgtaagt 32
Db 77822 TCATCTACTGAAACATGAGGAGGAAAGG 77852

RESULT 13
AC010593 93553 bp DNA linear HTG 18-JUL-2000
LOCUS Homo sapiens chromosome 19 clone CTC-47703, WORKING DRAFT SEQUENCE,
DEFINITION 8 ordered pieces.
AC010593
VERSION AC010593.4 GI:7711470
KEYWORDS HTG: HTGS_PHASE2; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 93553)
REFERENCE DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 19
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 93553)
REFERENCE DOE Joint Genome Institute.
TITLE Direct Submission
AUTHORS Submitted (16-SEP-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
JOURNAL On May 6, 2000 this sequence version replaced gi:7690058.
COMMENT -----
Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov

Project Information
Center Project Name: 368687, BC333711
Center clone name: CIT-HSPC_47703

Summary Statistics
Consensus quality: 89161 bases at least Q40
Consensus quality: 92004 bases at least Q30
Consensus quality: 92432 bases at least Q20
Estimated insert size: 99750; agarose-fp estimation
Estimated insert size: 93253; sum-of-contigs estimation
Quality coverage: 9.64 in Q20 bases; agarose-fp estimation
Quality coverage: 10.32 in Q20 bases; sum-of-contigs estimation.
NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1 2113: contig of 2113 bp in length
* 2114 2213: gap of unknown length

* 2214 44366: contig of 42153 bp in length
* 44367 44466: gap of unknown length
* 44467 63117: contig of 18651 bp in length
* 63118 63217: gap of unknown length
* 63218 64264: contig of 1047 bp in length
* 64265 64364: gap of unknown length
* 64365 64542: contig of 178 bp in length
* 64543 64642: gap of unknown length
* 64643 65631: contig of 989 bp in length
* 65632 65731: gap of unknown length
* 65732 86549: contig of 20818 bp in length
* 86550 93553: gap of unknown length
* 86550 Location/Qualifiers
1..93553
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/clone="CTC-47703"
/clone_lib="Caltech human BAC library C"
BASE COUNT 25617 a 19725 c 19529 g 27970 t 712 others
ORIGIN

Query Match 64.8%; Score 21.4; DB 2; Length 93553;
Best Local Similarity 80.6%; Pred. No. 79;
Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 2 tcatctactgcacagagagactgtaagt 32
Db 51464 TCATCTACTGAAACATGAGGAGGAAAGG 51434

RESULT 14
AC016587 155221 bp DNA linear PRI 21-JUL-2001
LOCUS Homo sapiens chromosome 19 clone CTD-2652B17, complete sequence.
DEFINITION AC016587
VERSION AC016587.8 GI:14993681
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 155221)
REFERENCE DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 155221)
REFERENCE DOE Joint Genome Institute.
TITLE Direct Submission
AUTHORS Submitted (04-DEC-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
JOURNAL 3 (bases 1 to 155221)
REFERENCE DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
AUTHORS Submitted (25-MAY-2001) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
JOURNAL 4 (bases 1 to 155221)
REFERENCE DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
AUTHORS Submitted (21-JUL-2001) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
JOURNAL On Jul 21, 2001 this sequence version replaced gi:14196366.
COMMENT Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 99.7% of Sequence;
Estimated Total Number of Errors is 0.3.
STS Content:
SHGC-5835 G14094.
Location/Qualifiers

source 1. .155221
/organism="Homo sapiens"

/db_xref="taxon:9606"
/chromosome="19"
/clone="CVD-2632B17"

BASE COUNT 41029 a 34040 c 33604 g 46548 t
ORIGIN

Query Match 64.8%; Score 21.4; DB 9; Length 155221;
Best Local Similarity 80.6%; Pred. No. 83;
Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 tcctactgcacagcgagactgttaagtgc 32
|||||
Db 149344 TCATCTACTGAAACATGAGAGGAAAGGG 149314

RESULT 15

LOCUS 163568 163568 234 bp DNA 1 linear PAT 07-OCT-1997

DEFINITION Sequence 143 from patent US 5663143.

ACCESSION 163568

VERSION 163568.1 GI:2481141

KEYWORDS

SOURCE

ORGANISM

Unclassified.

REFERENCE 1 (bases 1 to 234)

Levy, A., Charles, J., Ladner, R., Charles, J., Guterman, S., Kosow, N.,

Roberts, B., Lindsay, W., Markland, W., and Kent, R. Baripault.

Engineered human-derived kunitz domains that inhibit human

neutrophil elastase

Patent: US 5663143-A 143 02-SEP-1997;

Location/Qualifiers

1. .234

source /organism="unknown"

BASE COUNT 50 a 55 c 62 g 66 t 1 others

ORIGIN

Query Match 63.6%; Score 21; DB 6; Length 234;

Best Local Similarity 82.8%; Pred. No. 64;

Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 5 tctactgcacagcgagactgttaagtgc 33

|||||
Db 116 TCCTACTGCACTCGAGATGCTAAGTGC 144

Search completed: May 12, 2002, 20:13:05
Job time: 7167 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 12, 2002, 19:41:31 ; Search time 2487.21 Seconds
(without alignments)
341.872 Million cell updates/sec

Title: US-09-554-024-2
Perfect score: 1 ggtccaaagaagccagtgcc.....ctgtaagtcgacagagatg 63
Sequence: 1 ggtccaaagaagccagtgcc.....ctgtaagtcgacagagatg 63

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 674847542 residues
Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_estl:*
10: gb_est2:*
11: gb_hlc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pla:*
16: em_gss_vtl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	28.4	45.1	629	10	BF384319 602046667
2	28.4	45.1	825	10	BF234717 602028543
3	28.4	45.1	890	10	BF236794 602028092
4	27.4	43.5	652	9	A1573712 u164h04.y
5	26.8	42.5	426	10	W29802 mc24b05.r1
6	26.8	42.5	510	9	A1116411 u102906.y
7	26.8	42.5	560	9	A1037277 u606q11.y
8	26.8	42.5	585	10	BF138224 601781351
9	26.8	42.5	588	10	BF807277 2053-42 M
10	26.8	42.5	588	9	AM255323 ML332 pep
11	26.8	42.5	600	10	BG805295 0614-91 M
12	26.8	42.5	600	10	BG807215 2052-61 M
13	26.8	42.5	600	10	B1989219 4032-71 M
14	26.8	42.5	600	10	B1990353 4064-19 M
15	26.8	42.5	600	10	B1991233 4092-60 M
16	26.8	42.5	608	10	BG906827 Tair11548
17	26.8	42.5	627	9	AA871245 vq32h06.r

C 18	26.8	42.5	736	10	BI332864
C 19	26.8	42.5	754	10	BE915023
C 20	26.8	42.5	759	10	BI554490
C 21	26.8	42.5	774	9	AU080567
C 22	26.8	42.5	778	9	AU080499
C 23	26.8	42.5	822	10	BI833909
C 24	26.8	42.5	836	10	BI220094
C 25	26.8	42.5	910	9	A1266998
C 26	26.8	42.5	1100	10	BM461228
C 27	26.6	42.2	386	9	AA591138
C 28	26.6	42.2	552	12	AO516103
C 29	26.2	41.6	395	10	BG9696508
C 30	26	41.3	565	9	AA098372
C 31	26	41.3	986	10	BF781798
C 32	26	41.3	1708	10	BG685380
C 33	25.8	41.0	485	10	BG892285
C 34	25.8	41.0	595	9	A1195428
C 35	25.8	41.0	651	9	AM475350
C 36	25.2	40.0	496	10	BE496965
C 37	25.2	40.0	584	10	BF473423
C 38	25.2	40.0	600	10	BG803196
C 39	25.2	40.0	602	9	AA771450
C 40	25.2	40.0	661	10	BF570601
C 41	25.2	40.0	676	12	AG105074
C 42	25.2	40.0	715	10	BF578176
C 43	25.2	40.0	717	10	BE586077
C 44	25.2	40.0	760	9	A1182105
C 45	25.2	40.0	778	10	BI330345

ALIGNMENTS

RESULT 1
LOCUS BF384319/c
DEFINITION 602046667/1 NCL_CGAP_L19 Mus musculus cDNA clone IMAGE:419585 5',
ACCESSION BF384319
VERSION BF384319.1 GI:11365624
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 629)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LHAM9532 row: f column: 18
High quality sequence stop: 625.
location/Qualifiers
1. 629
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:419585"
/clone_lib="NCL_CGAP_L19"
/lab_host="DH10B (TI phage-resistant)"
/note="Organ: liver; Vector: pCMV-SPORT6; Site: 1; NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.9 kb. Constructed by Life
Technologies. Note: this is a NCL_CGAP Library."

BASE COUNT 169 a 144 c 174 g 142 t


```

/dev_stage="adult"
/lab_host="DH10B"
/note=Organ: liver; Vector: pME18S-FL3; Site_1: DraIII
(CACTGTGTC); Site_2: DraIII (CACCACGTGC); 1st strand cDNA
was primed with an oligo(dT) primer
[ATGGCGCCCTTTTTTTTTTTTTTTTT] double-stranded cDNA was
ligated to a DraIII adaptor [cgttcgaccttgcg] digested
and cloned into distinct DraIII sites of the pME18S-FL3
vector (5' site CACTGTGTC, 3' site CACCACGTGC). XhoI should
be used to isolate the cDNA insert. Size selection was
performed to exclude fragments <1.5kb. Library
constructed by Dr. Sumio Sugano (University of Tokyo
Institute of Medical Science). Custom primers for
sequencing: 5' end primer CTTCCTCTCTAAAGCTCCG and 3' end
primer CGAAGCTTGCAAGCTGAGCA."
BASE COUNT      128 a      118 c      133 g      131 t
ORIGIN

```

Query Match	Best Local Similarity	Best Match	Local Similarity	Score	DB	Length
37	68.5%	37	68.5%	26.8	DB 9	560
Conservative	0	Mismatches	17	Indels	0	Gaps
0	0	0	0	0	0	0

10 aagccagtgccaatcatctactgcacacgggagagcgtgtaagcagagagatg 63
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 510 AGGGCAGGGCCCTATTCTTCCCAATGAACATGACACATTGGAAAAATTCACAGGGGATG 457

BASE COUNT 157 a 122 c 145 g 136 t
 ORIGIN

was primed with an oligo(dT) primer
 [ATGTGGCCCTTTTTTTTTTTTTTTTTTTT]; double-stranded cDNA was
 ligated to a DraIII adaptor [TGTGGCCCTACTGG], digested
 and cloned into distinct draIII sites of the pME185-FL3
 vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should
 be used to isolate the cDNA insert. Size selection was
 performed to exclude fragments <1.5kb. Library
 constructed by Dr. Sumio Sugano (University of Tokyo
 Institute of Medical Science). Custom primers for 5' end
 sequencing: 5' end primer CTCTGCTCTTAAAGCTGCG and 3' end
 primer CGACCTGCGACCTCGAGACA."

DB	Accession	Source	Organism	Reference Authors	Title	Journal	Comment
db	328	AGGCGAGGCGCTATCTTCAAAATGACATGACGATTGGAAATTCACAGGGATG	275				
RESULT	7						
LOCUS	AI037277/c						
DEFINITION	AI037277	560 bp	mRNA	linear	EST 26-JUN-1998		
ACCESSION	U60911.y1	Sugano mouse liver	mlia	Mus musculus	CDNA clone		
VERSION	IMAGE:1450340.5	similar to gb:M74570	Mouse	aldhyde dehydrogenase			
KEYWORDS	II mRNA, complete cds (MOUSE);	mRNA sequence.					
SOURCE	AI037277	GI:3260968					
ORGANISM	EST.	house mouse.					
REFERENCE	Mus musculus						
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.						
	1 (bases 1 to 560)						
	Marré, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Thelasing, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.						
	The WashU-HHMI Mouse EST Project						
	Unpublished (1996)						
	Contact: Maria M/Mouse EST Project						
	WashU-HHMI Mouse EST Project						
	Washington University School of Medicine						
	4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108						
	Tel: 314 286 1800						
	Fax: 314 286 1810						
	Email: mouseest@watson.wustl.edu						
	This clone is available royalty-free through LNL: contact the						
	IMAGE Consortium (info@image.lnl.gov) for further information.						
	MGI:923656						
	Seq primer: custom primer used						
	High quality sequence stop: 417.						
	Location/Qualifiers						
	1. 560						
	/organism="Mus musculus"						
	/strain="C57BL"						
	/db_xref="taxon:10090"						
	/clone IMAGE:1450340"						
	/clone_11b="Sugano mouse liver mlia"						
	/sex="female"						
	/seq_stage="adult"						
	/lab_host="DH10B"						
	/note="Organ: liver; Vector: pME18S-FL3; Site_1: DraIII						
	(CACCCTGCTG); Site_2: DraIII (CACCATGCTG); 1st strand cDNA						

RESULT	8
BF138224/c	
LOCUS	BF138224 585 bp mRNA linear EST 24-OCT-2000
DEFINITION	601781351f1 NCI_CGAP_Lu30 Mus musculus cDNA clone IMAGE:4009333 5'
ACCESSION	BF138224
VERSION	BF138224.1 GI:10977264
KEYWORDS	EST.
SOURCE	house mouse.
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS	Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
TITLE	1 (bases 1 to 585)
JOURNAL	NIH-MGC http://mgc.nci.nih.gov/.
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgapds-remail.nih.gov Tissue Procurement: Gilbert Smith, Ph.D. CDNA Library Preparation: Life Technologies, Inc. CDNA library Arrayed by: The I.M.A.G.E. Consortium (LNLU) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov Plate: LHAM245 row: e column: 14 High quality sequence stop: 585. Location/Qualifiers 1..585 /organism="Mus musculus" /strain="CZECH II" /db_xref="taxon:10090" /clone="IMAGE:4009333" /clone_lib="NCI CGAP_Lu30" /tissue_type="tumor, metastatic to mammary" /lab_host="DH10B" /note="Organ: Lung; Vector: PCMV-SPORT6; Site_1: NotI; Site_2: SalI; transgenic model WNT-1, expression driven by MTV-LTR enhancer; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies. Investigator providing samples: Gilbert Smith, NIH"
BASE COUNT	147 a 136 c 160 g 142 t
ORIGIN	
Query Match	42.5%; Score 26.8; DB 10; Length 585;
Best Local Similarity	68.5%; Pred No. 1 3e+02;
Matches 37; Conservative	0; Mismatches 17; Indels 0; Gaps 0

[illegible]

FEATURES					
Source		Location/Qualifiers			
		1..589			
		/organism="Mentha x piperita"			
		/cultivar="Black Mitcham"			
		/db_xref="taxon:34256"			
		/clone_lib="peppermint glandular trichome"			
		/tissue_type="pellate glandular trichomes"			
		/cell_type="secretory"			
		/note="Vector: lambda ZAPIT"			
BASE COUNT		164 a 115 c 135 g 175 t			
ORIGIN					
<hr/>					
Query Match	42.5%	Score 26.8:	DB 9:	Length 589:	
Best Local Similarity	73.9%:	Pred. No. 1.3e+02:			
Matches 34:	Conservative	0:	Mismatches 12:	Indels	0:
Oy	9 gaagccagtgcgaatcactctactgcaaaaggagagactgttaagtcg 54				
Db	409 GAGGCCAGTTCTTCATCATCCACAGAACACACAGCAGCTGTAACAGC 364				
<hr/>					
RESULT 11					
BG805295/c					
LOCUS	BG805295	600 bp	mRNA	linear	EST 20-DEC-2001
DEFINITION	0614-91 Mouse E14.5 retina lambda ZAP II library Mus musculus CDNA,				
ACCESSION	BG805295				
VERSION	BG805295.1 GI:17952226				
KEYWORDS	EST.				
SOURCE	house mouse.				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
TITLE	1 (bases 1 to 600)				
JOURNAL	Mu, X., Zhao S., Pershad, R., Hsieh, T.-F., Scarpa, A., Wang, S.W.,				
COMMENT	White, R.A., Beremand, P.D., Thomas, T.L., Gan, L. and Klein, W.H. Gene expression in the developing mouse retina by EST sequencing and microarray analysis Nucleic Acids Res. 29 (24), 4983-4993 (2001) Contact: Klein WH Department of Biochemistry and Molecular Biology University of Texas M.D. Anderson Cancer Center Box 117, 1515 Holcombe Blvd., Houston, TX 77030, USA Tel: 713 792 3646 Fax: 713 790 0329.				
<hr/>					
FEATURES		Location/Qualifiers			
source		1..600			
		/organism="Mus musculus"			
		/db_xref="taxon:10090"			
		/clone_lib="Mouse E14.5 retina lambda ZAP II library"			
		/tissue_type="neural retina"			
		/dev_stage="embryonic day 14.5 post-fertilization"			
BASE COUNT		164 a 139 c 155 g 141 t 1 others			
ORIGIN					
<hr/>					
Query Match	42.5%	Score 26.8:	DB 10:	Length 600:	
Best Local Similarity	68.5%:	Pred. NO. 1.3e+02:			
Matches 37:	Conservative	0:	Mismatches 17:	Indels	0:
Oy	10 aaagcgatgcacatcatctactgcagaagggagctgaagtgcacaagatg 63				
Db	571 AGGGCAGGCGCCATATCTTCCAATGACATGAGCAATTGGAAATTTCCAGGGGATG 518				
<hr/>					
RESULT 12					
BG807215/c					
LOCUS	BG807215	600 bp	mRNA	linear	EST 20-DEC-2001
DEFINITION	2032-61 Mouse E14.5 retina lambda ZAP II library Mus musculus CDNA,				
ACCESSION	BG807215				
VERSION	BG807215.1 GI:17954177				

[illegible][illegible]

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 12, 2002, 18:12:43 ; Search time 2487.21 Seconds
(without alignments)
179.076 Million cell updates/sec

Title: US-09-554-024-1
Perfect score: 33
Sequence: 1 atcattctacgcacacagagagactgtaagtgc 33

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues
Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Length	DB ID	Description
C 1	23	69.7	647	9	AV708999 AV708999
C 2	21.4	64.8	215	10	BG062811 L0958812-
C 3	21.4	64.8	495	12	A0820935 HS_4679_B
C 4	21.4	64.8	1037	11	AK005310 Mus muscu
C 5	21	63.6	479	9	AU084050 AU084050
C 6	20.8	63.0	512	10	B1045689 MR3-FN020
C 7	20.8	63.0	1708	10	BG685380 602637225
C 8	20.4	61.8	344	10	B1142930 BK76d05.Y
C 9	20.4	61.8	346	10	BF721095 mab62h01.
C 10	20.4	61.8	382	9	A1153393 uc52e01.r
C 11	20.4	61.8	437	10	H29556 ym58d03.s1
C 12	20.4	61.8	470	9	AA071706 mm71a09.r
C 13	20.4	61.8	482	10	W78283 me78c12.r1
C 14	20.4	61.8	496	9	AA16356 ms71h03.r
C 15	20.4	61.8	529	10	BE631498 u60c08.Y
C 16	20.4	61.8	531	12	BH209770 Sml-47P15
C 17	20.4	61.8	560	9	A1047386 u65f09.Y

C 18	20.4	61.8	504	12	A2507444
C 19	20.4	61.8	753	10	B1149998
C 20	20.4	61.8	755	10	BF783942
C 21	20.4	61.8	786	10	BG870432
C 22	20.4	61.8	888	10	BF784120
C 23	20.4	61.8	2229	11	AK005184
C 24	20.2	61.2	312	10	BF737108
C 25	20.2	61.2	372	10	BG279687
C 26	20.2	61.2	396	10	BG96508
C 27	20.2	61.2	408	12	A0645225
C 28	20.2	61.2	486	12	A0642870
C 29	20.2	61.2	589	9	AM255323
C 30	20.2	61.2	636	12	A0948634
C 31	20.2	61.2	639	9	BB650920
C 32	20.2	61.2	665	10	BF480370
C 33	20.2	61.2	675	10	BM300861
C 34	20.2	61.2	694	10	BF479025
C 35	20.2	61.2	820	9	BE131115
C 36	20.2	61.2	472	10	B036406
C 37	20.2	60.6	521	9	AM128551
C 38	20.2	60.6	543	9	A1588657
C 39	20.2	60.6	574	9	AM078159
C 40	20.2	60.6	716	10	B1326776
C 41	20.2	60.6	860	10	BE900105
C 42	19.8	60.0	439	12	A2045606
C 43	19.8	60.0	484	12	A2730423
C 44	19.8	60.0	507	12	BH528457
C 45	19.8	60.0	534	9	AA719242

ALIGNMENTS

RESULT 1
LOCUS AV708999 647 bp mRNA linear EST 09-OCT-2000
DEFINITION AV708999
ACCESSION AV708999
VERSION AV708999.1 GI:10726264
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 647)
Peng,Y., Song,H., Huang,Q., Huang,C., Gu,Y., Yang,Y., Gao,G., Xiao
,H., Xu,X., Li,N., Qian,B., Liu,F., Ou,J., Gao,X., Cheng,Z., Xu,Z.,
Zeng,L., Xu,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu
,G., Hu,R., Chen,J., Chen,Z. and Han,Z.
Homo sapiens cDNA ADC clones
Unpublished (2000)
Contact: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.
Location/Qualifiers
1. 647
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_id="ADCAAF12"
/clone_lib="ADC"
/tissue_type="Adrenal gland"
/dev_stage="Adult"
/lab_host="SOLR"
/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"

BASE COUNT 136 a 172 c 176 g 162 t 1 others
ORIGIN

Query Match 69.7%; Score 23; DB 9; Length 647;
Best Local Similarity 83.9%; Pred. No. 95;
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 tcactctactgcaacagagagctgtaagt 32
|||||
Db 517 TCCTCTCTCAGACAGAGCATTTGGAGATG 487

RESULT 2
LOCUS BG062811/c 215 bp mRNA linear EST 25-JAN-2001
DEFINITION L0958B12-5 NIA Mouse Newborn Kidney cDNA library2 (Short) Mus
MUSCULUS CDNA clone L0958B12 5', mRNA sequence.
ACCESSION BG062811
VERSION BG062811.1 GI:12533662
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 215)
Plao,Y., Ko,N.T., Lim,M.K. and Ko,M.S.H.
Construction of long-transcript enriched cDNA libraries from
subclonable amounts of total RNAs by a universal PCR amplification
method
Genome Res. 11 (9), 1553-1558 (2001)
21429098

JOURNAL
MEDLINE
COMMENT
Contact: George J. Kargul
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: cdael@sun.gri.nih.gov
niast (http://lgsun.gri.nih.gov/cdna/cdna.html)
Plate: L0958 row: B column: 12
Seq primer: -21M13 Reverse
High quality sequence stop: 215
POLY-A-NO.

FEATURES
source
1. 215
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="niast:L0958B12-5"
/db_xref="taxon:10090"
/clone="L0958B12"
/clone_lib="NIA Mouse Newborn Kidney cDNA library2 (Short)"
/tissue_type="Newborn Kidney"
/dev_stage="Newborn"
/lab_host="DH10B"
/note="Vector: pSPORT1 (Invitrogen); Site_1: SalI; Site_2:
NotI; Mouse cDNA project by the Laboratory of Genetics,
National Institute on Aging (NIA), Intramural Research
Program, NIH (http://lgsun.gri.nih.gov/cdna). This is
a short-transcript enriched cDNA library (Ref. Genome Res.
11: 1553-1558 (2001). [PMID: 11544191]) In brief,
double-stranded cDNAs were synthesized with an oligo(dT)
primer (Invitrogen: 5'-
pACGAGTCCTGATCGCGAGCGCGCCCTTTTCTTTT-3') from 26
ug of total RNA, treated with T4 DNA polymerase, and
purified by ethanol-precipitation. The cDNAs were ligated
to lome-linker L1-SalI, purified by phenol/chloroform, and
separated from free linkers by Centricon 100. Then, the
cDNAs were amplified by long-range high fidelity PCR using
Ex Taq polymerase (Takara) with a primer SalI-4. The
products were purified by phenol/chloroform and Centricon
100. The cDNAs were digested with SalI and NotI enzymes
and cloned into SalI/NotI site of pSPORT1 plasmid vector.
The DH10B E. coli host was transformed with the ligation
mixture by the standard chemical method. The average
insert size is about 1.5 kb. The library was constructed
by Yulan Plao(NIA)."

BASE COUNT 73 a 46 c 44 g 52 t
ORIGIN

Query Match 64.8%; Score 21.4; DB 10; Length 215;
Best Local Similarity 80.6%; Pred. No. 2.8e+02;
Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 tcactctactgcaacagagagctgtaagt 32
|||||
Db 202 TCATCTACTGATCAGCAGACAGATGAGT 172

RESULT 3
LOCUS A0820935 495 bp DNA linear GSS 26-AUG-1999
DEFINITION HS_4679_B1_C07_SP6E CIT Approved Human Genomic Sperm Library D Homo
sapiens genomic clone Plate=4679 Col=13 Row=F, DNA sequence.
ACCESSION A0820935
VERSION A0820935.1 GI:5783328
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 495)
Keller,A., Shakeri,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
99380589

JOURNAL
MEDLINE
COMMENT
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones may be purchased from Research Genetics (Info@resgen.com).
BAC end Web Server: http://www.htsc.washington.edu
BAC: 4679 row: F column: 13
Seq primer: T7
Class: BAC ends
High quality sequence stop: 495.
Location/Qualifiers
1. 495
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=4679 Col=13 Row=F"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelobAC11; BAC clones in
E-Coli DH10B"

BASE COUNT 144 a 95 c 132 g 123 t 1 others
ORIGIN

Query Match 64.8%; Score 21.4; DB 12; Length 495;
Best Local Similarity 80.6%; Pred. No. 3.6e+02;
Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 atcactctactgcaacagagagctgtaagt 31
|||||
Db 241 ACCATCAACAGAAACAGAGAGTGAATT 271

RESULT 4
LOCUS AK005310 1037 bp mRNA linear HTC 19-JAN-2002
DEFINITION Mus musculus adult male cerebellum cDNA, RIKEN full-length enriched
library, clone:1500031D15; unclassified transcript, full insert

Query Match 63.6%: Score 21; DB 9; Length 479;
 Best Local Similarity 82.8%: Pred. No. 5.1e+02;
 Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 atcactctacgcacaggagctgta 29
 ||||| ||||| ||||| ||||| |||||
 Db 92 ATCAGCTACTTCACTGAGCCCTGATTA 120

RESULT 6
 BI045689 512 bp mRNA linear EST 14-JUN-2001
 LOCUS MR3-FN0206-300101-004-c09 FN0206 Homo sapiens CDNA, mRNA sequence.
 DEFINITION BI045689
 ACCESSION BI045689
 VERSION BI045689.1 GI:14452311
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS
 1. (bases 1 to 512)
 Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Brites,M.R.,
 Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
 Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
 Brunshtein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,
 M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
 Simpson,A.J.
 Shotgun sequencing of the human transcriptome with ORF expressed
 sequence tags
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 20202663
 COMMENT
 Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome
 Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?l1=MR3&t2=MR3-FN0206-
 300101-004-c09&t3=2001-01-30&t4=1)
 Seq primer: puc 18 forward
 High quality sequence start: 25
 High quality sequence stop: 510.
 Location/Qualifiers

FEATURES
 source
 1. 512
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="FN0206"
 /dev_stage="Adult"
 /note="Organ: prostate_normal; Vector: puc18; Site_1: SmaI
 ; Site_2: SmaI; A mini-library was made by cloning
 products derived from ORESTES PCR (U.S. Letters Patent
 application No. 196,716 - Ludwig Institute for Cancer
 Research) profiles into the puc 18 vector. Reverse
 transcription of tissue mRNA and cDNA amplification were
 performed under low stringency conditions."

BASE COUNT 166 a 144 c 114 g 88 t
 ORIGIN

Query Match 63.0%: Score 20.8; DB 10; Length 512;
 Best Local Similarity 78.1%: Pred. No. 6.2e+02;
 Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 1 atcactctacgcacaggagctgta 32
 ||||| ||||| ||||| ||||| |||||
 Db 461 ATCATCTAATGCACAGCAACAGACAGATG 492

RESULT 7

BG685380/c
 LOCUS 1708 bp mRNA linear EST 01-MAY-2001
 DEFINITION 602637225F1 NIH_MGC_48 Homo sapiens CDNA clone IMAGE:4764942 5',
 mRNA sequence.
 ACCESSION BG685380
 VERSION BG685380.1 GI:13916777
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS
 1. (bases 1 to 1708)
 NIH-MGC http://mgi.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 COMMENT
 Contact: Robert Strausberg, Ph.D.
 Email: cga@rsf@mail.nih.gov
 Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
 CDNA Library Preparation: Ling Hong/Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.llnl.gov
 Plate: LCM1623 row: a column: 07
 High quality sequence stop: 68.
 Location/Qualifiers

FEATURES
 source
 1. 1708
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="IMAGE:4764942"
 /clone_lib="NIH_MGC_48"
 /tissue_type="primary B-cells from tonsils (cell line)"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: B-cells; Vector: pORF7; Site_1: XhoI;
 Site_2: EcoRI; CDNA made by oligo-dT priming.
 Directionally cloned into EcoRI/XhoI sites using the
 following 5' adaptor: GGCAGGAG(G). Size-selected >500bp
 for average insert size 1.8kb. Library constructed by Ling
 Hong in the laboratory of Gerald M. Rubin (University of
 California, Berkeley) using ZAP-CDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies).
 Note: this is a NIH-MGC library."

BASE COUNT 486 a 456 c 394 g 372 t
 ORIGIN

Query Match 63.0%: Score 20.8; DB 10; Length 1708;
 Best Local Similarity 78.1%: Pred. No. 8.8e+02;
 Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 1 atcactctacgcacaggagctgta 32
 ||||| ||||| ||||| ||||| |||||
 Db 885 ATCATCTAATGCATAGTAGTACGATGAGTG 854

RESULT 8
 BI142930 344 bp mRNA linear EST 10-JUL-2001
 LOCUS FK76605.y1 Meloidogyne javanica egg PAMP1 v6 Chiapelli McCarter
 DEFINITION Meloidogyne javanica CDNA 5', mRNA sequence.
 ACCESSION BI142930
 VERSION BI142930.1 GI:14624640
 KEYWORDS EST.
 SOURCE root-knot nematode.
 ORGANISM Meloidogyne javanica

REFERENCE
 AUTHORS
 1. (bases 1 to 344)
 McCarter,J., Clifton,S., Chiapelli,B., Pape,D., Martin,J., Wylie,T.,
 Dante,M., Marra,M., Hillier,L., Kucaba,T., Theising,B., Bowers,Y.,
 Gibbons,M., Rutter,E., Bennett,J., Franklin,C., Tsagarisvill,R.,
 Ronko,I., Kennedy,S., Maguire,L., Beck,C., Underwood,K., Steptoe

BASE COUNT	109 a	74 c	84 g	115 t	
ORIGIN					
Query Match	61.88;	Score 20.4;	DB 9;	Length 382;	
Best Local Similarity	80.0%;	Pred. No. 8.1e+02;			
Matches 24;	Conservative 0;	Mismatches 6;	Indels 0;	Gaps 0;	
Oy	3	catctactgcaacagagagactgtgaagtg 32			
Db	173	CATCTTACTCCAGCAGCAGCAGCAGCTGATG 144			
RESULT 11					
129556/c					
LOCUS					
DEFINITION	H29556 437 bp mRNA linear EST 17-JUL-1995				
ACCESSION	YMS6d03.s1 Soares infant brain INIB Homo sapiens cDNA clone				
VERSION	IMAGE:5280 3' similar to contains Alu repetitive element;contains				
KEYWORDS	LTR3 Repetitive element ;, mRNA sequence.				
SOURCE	H29556 H29556.1 GI:900466				
ORGANISM	EST.				
	human.				
	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
REFERENCE	1 (bases 1 to 437)				
AUTHORS	Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman				
	,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,				
	Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston				
	,R., Williamson,A., Woldmann,P. and Wilson,R.				
TITLE	The Washu-Merck EST Project				
JOURNAL	Unpublished (1995)				
COMMENT	Contact: Wilson RK				
	Washington University School of Medicine				
	4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108				
	Tel: 314 266 1800				
	Fax: 314 266 1810				
	Email: estelw@wustl.edu				
	Insert Size: 1441				
	High quality sequence stops: 355 Source: IMAGE Consortium, LNL				
	This clone is available royalty-free through LNL ; contact the				
	IMAGE Consortium (info@image.lnl.gov) for further information.				
	Insert Length: 1441 Std Error: 0.00				
	Seq primer: Promega -21m13				
	High quality sequence stop: 355.				
FEATURES	Location/Qualifiers				
SOURCE	1. 437				
	/organism="Homo sapiens"				
	/db_xref="GDB:425816"				
	/db_xref="taxon:9606"				
	/clone="IMAGE:5280"				
	/clone_lib="Soares infant brain INIB"				
	/sex="Female"				
	/dev_stage="73 days post natal"				
	/lab_host="DH10B (ampicillin resistant)"				
	/note="Organ: whole brain; Vector: Lambda BA; Site:1: Not				
	I; Site:2: Hind III; 1st strand cDNA was primed with a Not				
	I - oligo(dT) primer [5'				
	AAGTGGAGAAATTCGGCCGCGCAGGAATTTTTTTTTTTTTTTT 3'];				

Query Match	61.8%	Score 20.4	DB 10:	Length 437;
Best Local Similarity	80.0%	Pred. No. 8.4e+02;		
Matches 24;	Conservative 0;	Mismatches 6;	Indels 0;	Gaps 0;
QY 2	tcactactgcacagagagactgtaagt 31			
Db 379	TCCTATCCAGAAACAGAGAACTGCTAGT 350			
RESULT 12	AA071706	470 bp	mRNA	linear
LOCUS	AA071706/C	mm71a09.r1	Stratagene mouse macrophage (#937306)	Mus musculus CDNA
DEFINITION	clone IMAGE:533848 5', mRNA sequence.			
ACCESSION	AA071706			
KEYWORDS	AA071706.1	GI:1590052		
VERSION	EST.			
ORGANISM	mouse mouse.			
SOURCE	Mus musculus			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 470)			
AUTHORS	Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisler,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Scheinlenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.			
TITLE	The WashU-HMIT Mouse EST Project			
JOURNAL	Unpublished (1996)			
COMMENT	Contact: Marra M/Mouse EST Project WashU-HMIT Mouse EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@watson.wustl.edu This clone is available royalty-free through LNL; contact the IMAGE Consortium (infoimage.llnl.gov) for further information. MGI:320784 Seq primer: -28ml3 rev1 ET from Amersham High quality sequence stop: 444. Location/Qualifiers 1..470 /organism="Mus musculus" /db_xref="taxon:10090" /clone_image="IMAGE:533848" /clone_lib="Stratagene mouse macrophage (#937306)" /tissue_type="macrophage" /dev_stage="MEH1-3 cell line" /lab_host="SOLR (kanamycin resistant)" /note="Organ: blood; Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo dT. MEH1-3 cell line. Average insert size: 1.5 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAAATCGCGACGAG 3' -3' adaptor sequence: 5' CTCGAGTTTATTTTATTTT 3'"			
BASE COUNT	135 a	86 c	105 g	144 t
ORIGIN				
Query Match	61.8%	Score 20.4	DB 9:	Length 470;
Best Local Similarity	80.0%	Pred. No. 8.6e+02;		
Matches 24;	Conservative 0;	Mismatches 6;	Indels 0;	Gaps 0;
OY 3	catctactgcacacagagagactgtaagt 32			

Db 149 CATCTACTCCAGCAGCAGCAGTGAGTG 120
 RESULT 13
 LOCUS W78283 482 bp mRNA linear EST 21-JUN-1996
 DEFINITION me78c12.r1 Soares mouse embryo NBME13.5 14.5 Mus musculus cDNA
 accession W78283
 version W78283.1 GI:1388746
 keywords EST.
 source house mouse.
 organism Mus musculus
 reference Eukaryota: Chordata: Craniata: Vertebrata: Euteleostomi: Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus.
 authors Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Scheilenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
 title The WashU-HMI Mouse EST Project
 journal Unpublished (1996)
 comment Contact: Marra M/Mouse EST Project
 WashU-HMI Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@watson.wustl.edu
 This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:245454
 seq primer: EPPRimer
 High quality sequence stop: 267.
 location/qualifiers
 1.482
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="IMAGE:401686"
 /clone_lib="Soares mouse embryo NBME13.5 14.5"
 /sex="unknown"
 /tissue_type="embryo"
 /dev_stage="13.5-14.5dpc total fetus"
 /lab_host="DH10B"
 /note="Vector: pT773D-Pac (Pharmacia) with a modified polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTACCAATCTGACGTGGAGCGCGCGGAAATTTTTTTTTTTTTTTTTTTT 3'], on equal amounts of mRNA from 2 13.5dpc and 2 14.5dpc embryos [total RNA provided by Minoru Ko, Wayne State Univ., from 2]; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT773 vector. Library went through one round of normalization, and was constructed by Bento Soares and M.Fatima Bonaldo."
 BASE COUNT 152 a 98 c 113 g 119 t
 ORIGIN
 Query Match 61.8%; Score 20.4; DB 10; Length 482;
 Best Local Similarity 80.0%; Pred. No. 8.6e+02;
 Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
 Oy 3 catctactgcaacagagagctgtaagt 32
 Db 292 CATCTACTCCAGCAGCAGCAGTGAGTG 263
 RESULT 14

AA166356/c
 LOCUS AA166356 496 bp mRNA linear EST 16-FEB-1997
 DEFINITION ms71n03.r1 Soares mouse 3NBMS Mus musculus cDNA clone IMAGE:617045
 accession AA166356
 version AA166356.1 GI:1744447
 keywords EST.
 source house mouse.
 organism Mus musculus
 reference Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi: Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus.
 authors Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Scheilenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
 title The WashU-HMI Mouse EST Project
 journal Unpublished (1996)
 comment Contact: Marra M/Mouse EST Project
 WashU-HMI Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@watson.wustl.edu
 This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:377869
 seq primer: -28M13 rev2 from Amersham
 High quality sequence stop: 333.
 location/qualifiers
 1.496
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="IMAGE:617045"
 /clone_lib="Soares mouse 3NBMS"
 /sex="male"
 /tissue_type="Spleen"
 /dev_stage="4 weeks"
 /lab_host="DH10B"
 /note="Vector: pT773D-Pac (Pharmacia) with a modified polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTACCAATCTGACGTGGAGCGCGCGCGGAAATTTTTTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT773 vector. RNA provided by Dr. Bertrand Jordan. Library went through three rounds of normalization, and was constructed by Bento Soares and M.Fatima Bonaldo."
 BASE COUNT 146 a 95 c 108 g 147 t
 ORIGIN
 Query Match 61.8%; Score 20.4; DB 9; Length 496;
 Best Local Similarity 80.0%; Pred. No. 8.7e+02;
 Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
 Oy 3 catctactgcaacagagagctgtaagt 32
 Db 201 CATCTACTCCAGCAGCAGCAGTGAGTG 172
 RESULT 15
 BE631498/c
 LOCUS BE631498 529 bp mRNA linear EST 25-AUG-2000
 DEFINITION uu60c08.y1 Soares_thymus_2NBMT Mus musculus cDNA clone
 accession BE631498
 version BE631498.1 GI:9914186
 keywords EST.

SOURCE
ORGANISM

house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 529)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

AUTHORS
TITLE

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

JOURNAL
COMMENT

Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:1085938
Seq primer: -40RP from Gibco
High quality sequence stop: 467.
Location/Qualifiers

FEATURES
SOURCE

1..529
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_image="3376334"
/clone_lib="Soares_Chymus_2NbMT"
/sex="male"
/tissue_type="Thymus"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Vector: pT7r3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTTCACCAATCTGAGAGTGGAGCGCGCGTGTGTGTGTGTGTGTGTGT
3']; double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7r3 vector. RNA
provided by Dr. Bertrand Jordan. Library went through two
rounds of normalization, and was constructed by Bento
Soares and M.Felima Bonaldo."

BASE COUNT
ORIGIN

147 a 102 c 122 g 158 t

Query Match 61.8%; Score 20.4; DB 10; Length 529;
Best Local Similarity 80.0%; Pred. No. 8.9e+02;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Oy 3 catctactgcaacagcagagactgtaagt 32
||||||| ||||| ||||| |||||
Db 173 CATCTACTCCAGCAGCGACAGAGGTGACTG 144

Search completed: May 12, 2002, 19:41:31
Job time: 5328 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OW nucleic - nucleic search, using sw model

Run on: May 12, 2002, 19:41:44 ; Search time 96.6 Seconds
(without alignments)

133.857 Million cell updates/sec

Title: US-09-554-024-1

Perfect score: 33
Sequence: 1 atcatctactgcacacagagagactgtaagtgc 33

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 398110 seqs, 195917913 residues

Total number of hits satisfying chosen parameters: 796220

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Pending_Patents_NA_New:*
1: /cgn2_6/ptodata/1/pna/PCF_NEW_COMB.seq:*
2: /cgn2_6/ptodata/1/pna/US06_NEW_COMB.seq:*
3: /cgn2_6/ptodata/1/pna/US07_NEW_COMB.seq:*
4: /cgn2_6/ptodata/1/pna/US08_NEW_COMB.seq:*
5: /cgn2_6/ptodata/1/pna/US09_NEW_COMB.seq:*
6: /cgn2_6/ptodata/1/pna/US10_NEW_COMB.seq:*
7: /cgn2_6/ptodata/1/pna/US60_NEW_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	19.4	58.8	16463	6	US-10-105-299-10527
C 2	19.2	58.2	5053	6	US-10-109-310-20
C 3	19.2	58.2	33000	6	US-10-109-310-18
C 4	18.6	56.4	588	5	US-09-749-7288-4
C 5	18.6	56.4	2005	5	US-09-636-596C-18
C 6	18.6	56.4	9792	5	US-09-636-596C-14
C 7	18.6	56.4	14155	6	US-10-108-605-102
C 8	18.6	56.4	49642	6	US-10-105-299-10868
C 9	18.4	55.8	1511	5	US-09-612-402A-10
C 10	18.4	55.8	1515	5	US-09-612-402A-22
C 11	18.4	55.8	3324	5	US-09-612-402A-24
C 12	18.4	55.8	3354	5	US-09-612-402A-23
C 13	18.4	55.8	4435	5	US-09-612-402A-1
C 14	18.2	55.2	34547	6	US-10-105-299-7482
C 15	18.2	55.2	224	6	US-10-113-872-1613
C 16	18.2	55.2	275	6	US-10-015-219-710
C 17	18.2	55.2	275	6	US-10-015-219-1447
C 18	18.2	55.2	282	6	US-10-015-219-1091
C 19	18.2	55.2	282	6	US-10-015-219-1441
C 20	18.2	55.2	291	6	US-10-015-219-1186
C 21	18.2	55.2	291	6	US-10-015-219-1450
C 22	18.2	55.2	2650	6	US-10-106-698-2090
C 23	18.2	55.2	3347	6	US-10-000-773A-16
C 24	18.2	55.2	6739	1	PCR-US02-08253-250
C 25	18.2	55.2	6739	1	US-10-116-802-237
C 26	17.8	53.9	96	5	US-09-539-331D-23554

C 27	17.8	53.9	98	5	US-09-540-210B-6858	Sequence 6858, Ap
C 28	17.8	53.9	153	5	US-09-540-210B-29815	Sequence 29815, A
C 29	17.8	53.9	407	6	US-10-106-698-2497	Sequence 2497, Ap
C 30	17.8	53.9	460	5	US-09-539-331D-35624	Sequence 35624, A
C 31	17.8	53.9	1723	6	US-10-105-299-1536	Sequence 1536, Ap
C 32	17.8	53.9	2704	6	US-10-116-512-3	Sequence 3, Appl
C 33	17.8	53.9	3179	1	PCR-US02-09135-134	Sequence 134, Appl
C 34	17.8	53.9	3179	1	PCR-US02-09105-191	Sequence 191, Appl
C 35	17.8	53.9	3179	1	PCR-US02-09188-390	Sequence 390, Appl
C 36	17.8	53.9	3179	1	PCR-US02-09257-257	Sequence 257, Appl
C 37	17.8	53.9	3179	1	PCR-US02-09370-404	Sequence 404, Appl
C 38	17.8	53.9	3179	1	PCR-US02-09922-253	Sequence 253, Appl
C 39	17.8	53.9	3179	6	US-10-105-299-3099	Sequence 3099, Ap
C 40	17.8	53.9	3258	5	US-09-540-209B-8	Sequence 8, Appl
C 41	17.8	53.9	4015	6	US-10-106-698-1018	Sequence 1018, Ap
C 42	17.8	53.9	4852	6	US-10-105-299-10103	Sequence 10103, A
C 43	17.6	53.3	151	5	US-09-685-209A-68	Sequence 68, Appl
C 44	17.6	53.3	168	5	US-09-975-254-28871	Sequence 28871, A
C 45	17.6	53.3	223	5	US-09-540-210B-24915	Sequence 24915, A

ALIGNMENTS

RESULT 1
US-10-105-299-10527/c
Sequence 10527, Application US/10105299
GENERAL INFORMATION:
APPLICANT: Rosen, et. al
TITLE OF INVENTION: Human Secreted Proteins
FILE REFERENCE: PS950
CURRENT APPLICATION NUMBER: US/10/105,299
CURRENT FILING DATE: 2002-03-26
NUMBER OF SEQ ID NOS: 15197
Prior Application removed - See File Wrapper or Palm
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 10527
LENGTH: 16463
TYPE: DNA
ORGANISM: Homo sapiens
US-10-105-299-10527

Query Match 58.8% Score 19.4; DB 6; Length 16463;
Best Local Similarity 79.3% Pred. No. 20;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 atcatctactgcacacagagagactgtaagtgc 29
Db 10709 ATCATCCACTGTTACAGAGGCTGTGAA 10681

RESULT 2
US-10-109-310-20
Sequence 20, Application US/10109310
GENERAL INFORMATION:
APPLICANT: Wisconsin Alumni Research Foundation
APPLICANT: Hutchinson, Charles R.
APPLICANT: Kennedy, Jonathan n.m.1
APPLICANT: Park, Cheonsok n.m.1
TITLE OF INVENTION: METHOD OF PRODUCING ANTIHYPERCHOLESTEROLEMIC AGENTS
FILE REFERENCE: 960296.95718
CURRENT APPLICATION NUMBER: US/10/109,310
CURRENT FILING DATE: 2002-03-28
PRIOR APPLICATION NUMBER: US/09/215,694
PRIOR FILING DATE: 1999-12-18
NUMBER OF SEQ ID NOS: 36
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 20
LENGTH: 5053
TYPE: DNA
ORGANISM: Aspergillus terreus
US-10-109-310-20

Query Match 58.2%; Score 19.2; DB 6; Length 5053;
Best Local Similarity 75.0%; Pred. No. 20;
Matches 24; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
Qy 1 atcatctactgcacagaggaactgtaagtgc 32
|||||
Db 51 atcatctactgcacagaggaactgtaagtgc 82

RESULT 3
US-10-109-310-18/C
; Sequence 18, Application US/10109310
; GENERAL INFORMATION:
; APPLICANT: Wisconsin Alumni Research Foundation
; APPLICANT: Hutchinson, Charles R.
; APPLICANT: Kennedy, Jonathan n.m.1
; APPLICANT: Park, Cheonsaek n.m.1
; TITLE OF INVENTION: METHOD OF PRODUCING ANTIHYPERCHOLESTEROLEMIC AGENTS
; FILE REFERENCE: 960296, 95718
; CURRENT APPLICATION NUMBER: US/10/109,310
; CURRENT FILING DATE: 2002-03-28
; PRIOR APPLICATION NUMBER: US/09/215,694
; PRIOR FILING DATE: 1999-12-18
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 18
; LENGTH: 33000
; TYPE: DNA
; ORGANISM: Aspergillus terreus
; US-10-109-310-18

Query Match 58.2%; Score 19.2; DB 6; Length 33000;
Best Local Similarity 75.0%; Pred. No. 28;
Matches 24; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 1 atcatctactgcacagaggaactgtaagtgc 32
|||||
Db 6815 atcatctactgcacagaggaactgtaagtgc 6784

RESULT 4
US-09-749-728B-4
; Sequence 4, Application US/09749728B
; GENERAL INFORMATION:
; APPLICANT: Umezawa, Akihiko
; APPLICANT: Hata, Jun-ichi
; APPLICANT: Fukuda, Keiichi
; APPLICANT: Ogawa, Satoshi
; APPLICANT: Sakurada, Kazuhiko
; APPLICANT: Gojo, Satoshi
; APPLICANT: Yamada, Yoji
; TITLE OF INVENTION: THE CELL HAVING THE POTENTIALITY OF DIFFERENTIATION INTO CARDIOMYOCYTES
; FILE REFERENCE: 00766, 000043
; CURRENT APPLICATION NUMBER: US/09/749,728B
; CURRENT FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: H11-372826
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: PCT-JP00-01148
; PRIOR FILING DATE: 2000-02-28
; PRIOR APPLICATION NUMBER: PCT-JP00-07741
; PRIOR FILING DATE: 2000-11-02
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 4
; LENGTH: 588
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; OTHER INFORMATION: (1)..(591)

US-09-749-728B-4

Query Match 56.4%; Score 18.6; DB 5; Length 588;
Best Local Similarity 72.7%; Pred. No. 26;
Matches 24; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
Qy 1 atcatctactgcacagaggaactgtaagtgc 33
|||
Db 388 accgcgtctgcacacagaggaactgtaagtgc 420

RESULT 5
US-09-636-596C-18
; Sequence 18, Application US/09636596C
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM
; TITLE OF INVENTION: PHOSPHODIESTER ALPHA-GLUCANASE OF THE LYSOSOMAL TARGETING PATH
; FILE REFERENCE: 10929-0001-77
; CURRENT APPLICATION NUMBER: US/09/636,596C
; CURRENT FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: Patentln version 3.0
; SEQ ID NO 18
; LENGTH: 2005
; TYPE: DNA
; ORGANISM: Mus musculus
; US-09-636-596C-18

Query Match 56.4%; Score 18.6; DB 5; Length 2005;
Best Local Similarity 72.7%; Pred. No. 32;
Matches 24; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 1 atcatctactgcacagaggaactgtaagtgc 33
|||
Db 1394 atcaccactgcagaggtgaacgaggaagcgc 1426

RESULT 6
US-09-636-596C-14
; Sequence 14, Application US/09636596C
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM
; TITLE OF INVENTION: PHOSPHODIESTER ALPHA-GLUCANASE OF THE LYSOSOMAL TARGETING PATH
; FILE REFERENCE: 10929-0001-77
; CURRENT APPLICATION NUMBER: US/09/636,596C
; CURRENT FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: Patentln version 3.0
; SEQ ID NO 14
; LENGTH: 9792
; TYPE: DNA
; ORGANISM: Mus musculus
; US-09-636-596C-14

Query Match 56.4%; Score 18.6; DB 5; Length 9792;
Best Local Similarity 72.7%; Pred. No. 42;
Matches 24; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 1 atcatctactgcacagaggaactgtaagtgc 33
|||
Db 9181 atcaccactgcagaggtgaacgaggaagcgc 9213

RESULT 7
US-10-108-605-102
; Sequence 102, Application US/10108605

```

? GENERAL INFORMATION:
? APPLICANT: Broadus, Julie
? APPLICANT: Stem, Lynn
? APPLICANT: Bachmann, Jane
? APPLICANT: Kandari, Kim
? TITLE OF INVENTION: NUCLEIC ACID SEQUENCES FROM DROSOPHILA MELANOGASTER THAT ENCODE
? TITLE OF INVENTION: PROTEINS ESSENTIAL FOR LARVAL VIABILITY AND USES THEREOF
? FILE REFERENCE: 3113B
? CURRENT APPLICATION NUMBER: US/10/108,605
? CURRENT FILING DATE: 2002-03-27
? PRIOR APPLICATION NUMBER: US 09/761,142
? PRIOR FILING DATE: 2001-01-16
? PRIOR APPLICATION NUMBER: US 60/176,418
? PRIOR FILING DATE: 2000-01-14
? NUMBER OF SEQ ID NOS: 361
? SOFTWARE: PatentIn Ver. 2.1
? SEQ ID NO 102
? LENGTH: 14155
? TYPE: DNA
? ORGANISM: Drosophila melanogaster
? US-10-108-605-102

```

Query Match	56.4%	Score 18.6	DB 6	Length 14155
Best Local Similarity	72.7%	Pred. No. 45		
Matches 24	Conservative 0	Mismatches 9	Indels 0	Gaps 0
QY	1	atcatctactgcacaacgagagactggtgaatgac	33	
b	3548	accacaagactgcacgtgtacacactgtgagatgc	3580	

```

US-10-105-299-10868
RESULT 8
US-10-105-299-10868
; Sequence 10868, Application US/10105299
; GENERAL INFORMATION:
; APPLICANT: Rosen, et. al
; TITLE OF INVENTION: Human Secreted Proteins
; FILE REFERENCE: P9350
; CURRENT APPLICATION NUMBER: US/10/105,299
; CURRENT FILING DATE: 2002-03-26
; NUMBER OF SEQ ID NOS: 15197
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10868
; LENGTH: 49642
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-105-299-10868

```

Query March	56.4%	Score 18.6	DB 6	Length 49642
Best Local Similarity	84.0%	Pred. No. 56		
Matches 21: Conservative	0	Mismatches	4	Indels 0; Gaps 0
OY	1	atcatctactctgcaacagagagactg	25	
db	23510	atcatctactctgtagtagagagactg	23534	

```

RESULT      9
US-09-612-402A-10
; Sequence 10, Application US/09612402A
; GENERAL INFORMATION:
; APPLICANT: Jackson, W. James
; APPLICANT: Pace, John
; TITLE OF INVENTION: Chlamydia Protein, Gene Sequence and Uses Thereof
; FILE REFERENCE: 7969-086-999
; CURRENT APPLICATION NUMBER: US/09/612,402A
; CURRENT FILING DATE: 2002-03-29
; PRIOR APPLICATION NUMBER: 08/942,596
; PRIOR FILING DATE: 1997-10-02
; NUMBER OF SEQ ID NOS: 43

```

```

; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 10
; LENGTH: 1511
; TYPE: DNA
; ORGANISM: Chlamydia sp.
US-09-612-402A-10

```

Query Match	55.8%;	Score 18.4;	DB 5;	Length 1511;
Similarity	78.6%;	Pred. No. 37;		
Best Local				
Matches	22;	Mismatches	6;	Indels 0;
Conservative				Gaps 0;

```

QY      1 atcatctactgcacacagagactgta 28
          |||||  |||||  |  |||
Db      696 atcatctacttcacacagaagatccagta 723

```

```

RESULT 10
US-09-612-402A-22
: Sequence 22, Application US/09612402A
: GENERAL INFORMATION:
: APPLICANT: Jackson, W. James
: APPLICANT: Pace, John
: TITLE OF INVENTION: Chlamydia Protein, Gene Sequence and Uses Thereof
: FILE REFERENCE: 7969-086-999
: CURRENT APPLICATION NUMBER: US/09/612,402A
: CURRENT FILING DATE: 2002-03-29
: PRIOR APPLICATION NUMBER: 08/942,596
: PRIOR FILING DATE: 1997-10-02
: NUMBER OF SEQ ID NOS: 43
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 22
: LENGTH: 1515
: TYPE: DNA
: ORGANISM: Chlamydia sp.
US-09-612-402A-22

```

Query Match	55.8%;	Score	18.4;	DB	5;	Length	1515;
Best Local Similarity	78.6%;	Pred. No.	37;				
Matches	22;	Mismatches	6;	Indels	0;	Gaps	0;
		Conservative					

Qy 1 atcatctactgcaacaggaagactgta 28
 ||||| ||||| | |||
 Db 696 atcatctacttcaacagaagaatccagta 723

```

RESULT 11
US-09-612-402A-24
: Sequence 24, Application US/09612402A
: GENERAL INFORMATION:
: APPLICANT: Jackson, W. James
: TITLE OF INVENTION: Chlamydia Protein, Gene Sequence and Uses Thereof
: FILE REFERENCE: 7969-086-999
: CURRENT APPLICATION NUMBER: US/09/612,402A
: PRIOR FILING DATE: 2002-03-29
: PRIOR APPLICATION NUMBER: 08/942,596
: PRIOR FILING DATE: 1997-10-02
: NUMBER OF SEQ ID NOS: 43
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 24
: LENGTH: 3324
: TYPE: DNA
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Description of Artificial Sequence: Recombinant Expression Vec
US-09-612-402A-24

```

Query Match	55.8%	Score 18.4	DB 5	Length 3324
Best Local Similarity	78.6%	Pred. No. 43		
Matches 22: Conservative	0	Mismatches 6	Indels 0	Gaps 0

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 12, 2002, 18:59:44 ; Search time 4348.96 seconds

(without alignments)
164.166 Million cell updates/sec

Title: US-09-554-024-1

Perfect score: 33
1 atcattctctgcacagagagactgtaagtgc 33

Scoring table:
IDENTITY-NUC
Gapop 10.0 , Gapext 1.0

Searched: 21979536 seqs, 10817449327 residues

Total number of hits satisfying chosen parameters: 43959072

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : Pending_Patents_NA_Main:*

1: /cgn2_6/ptodata/2/pna/US0506_COMB.seq:*
2: /cgn2_6/ptodata/2/pna/US0506_COMB.seq:*
3: /cgn2_6/ptodata/2/pna/US07_COMB.seq:*
4: /cgn2_6/ptodata/2/pna/US080_COMB.seq:*
5: /cgn2_6/ptodata/2/pna/US081_COMB.seq:*
6: /cgn2_6/ptodata/2/pna/US082_COMB.seq:*
7: /cgn2_6/ptodata/2/pna/US083_COMB.seq:*
8: /cgn2_6/ptodata/2/pna/US084_COMB.seq:*
9: /cgn2_6/ptodata/2/pna/US085_COMB.seq:*
10: /cgn2_6/ptodata/2/pna/US086_COMB.seq:*
11: /cgn2_6/ptodata/2/pna/US087_COMB.seq:*
12: /cgn2_6/ptodata/2/pna/US088_COMB.seq:*
13: /cgn2_6/ptodata/2/pna/US089_COMB.seq:*
14: /cgn2_6/ptodata/2/pna/US090_COMB.seq:*
15: /cgn2_6/ptodata/2/pna/US091_COMB.seq:*
16: /cgn2_6/ptodata/2/pna/US092_COMB.seq:*
17: /cgn2_6/ptodata/2/pna/US093_COMB.seq:*
18: /cgn2_6/ptodata/2/pna/US094_COMB.seq:*
19: /cgn2_6/ptodata/2/pna/US095_COMB.seq:*
20: /cgn2_6/ptodata/2/pna/US095A_COMB.seq:*
21: /cgn2_6/ptodata/2/pna/US095C_COMB.seq:*
22: /cgn2_6/ptodata/2/pna/US095D_COMB.seq:*
23: /cgn2_6/ptodata/2/pna/US095E_COMB.seq:*
24: /cgn2_6/ptodata/2/pna/US095B_COMB.seq:*
25: /cgn2_6/ptodata/2/pna/US096_COMB.seq:*
26: /cgn2_6/ptodata/2/pna/US096C_COMB.seq:*
27: /cgn2_6/ptodata/2/pna/US096D_COMB.seq:*
28: /cgn2_6/ptodata/2/pna/US096E_COMB.seq:*
29: /cgn2_6/ptodata/2/pna/US097A_COMB.seq:*
30: /cgn2_6/ptodata/2/pna/US097B_COMB.seq:*
31: /cgn2_6/ptodata/2/pna/US097C_COMB.seq:*
32: /cgn2_6/ptodata/2/pna/US098A_COMB.seq:*
33: /cgn2_6/ptodata/2/pna/US098B_COMB.seq:*
34: /cgn2_6/ptodata/2/pna/US098C_COMB.seq:*
35: /cgn2_6/ptodata/2/pna/US099A_COMB.seq:*
36: /cgn2_6/ptodata/2/pna/US099B_COMB.seq:*
37: /cgn2_6/ptodata/2/pna/US099C_COMB.seq:*
38: /cgn2_6/ptodata/2/pna/US100_COMB.seq:*
39: /cgn2_6/ptodata/2/pna/US101_COMB.seq:*
40: /cgn2_6/ptodata/2/pna/US6000_COMB.seq:*
41: /cgn2_6/ptodata/2/pna/US6001_COMB.seq:*
42: /cgn2_6/ptodata/2/pna/US6002_COMB.seq:*
43: /cgn2_6/ptodata/2/pna/US6003_COMB.seq:*
44: /cgn2_6/ptodata/2/pna/US6004_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	# Match	Query Length	ID	Description
1	33	100.0	33	US-09-554-024-1	Sequence 1, App1
2	33	100.0	63	US-09-554-024-2	Sequence 2, App1
3	33	100.0	87	US-09-554-024-9	Sequence 9, App1
4	33	100.0	98	US-09-554-024-3	Sequence 3, App1
5	33	100.0	197	US-09-554-024-5	Sequence 5, App1
6	22.4	67.9	3102	US-08-265-534A-5	Sequence 5, App1
7	22.4	67.9	3102	US-08-456-341-5	Sequence 5, App1
8	22.4	67.9	3102	US-08-457-001-5	Sequence 5, App1
9	22.4	67.9	3102	US-08-457-349-5	Sequence 5, App1
10	22.4	67.9	3102	US-08-457-543-5	Sequence 5, App1
11	22	66.7	1230	US-09-248-796-8333	Sequence 8333, Ap
12	22	66.7	1230	US-09-248-796-8333	Sequence 8333, Ap
13	21.4	64.8	12414	US-09-620-392-67364	Sequence 35280, A
14	21.4	64.8	12414	US-09-620-392-67364	Sequence 35280, A
15	21.4	64.8	12414	US-09-620-392-67364	Sequence 35280, A
16	21.4	64.8	12414	US-09-620-392-67364	Sequence 35280, A
17	21.4	64.8	12414	US-09-620-392-67364	Sequence 35280, A
18	21.4	64.8	12414	US-09-620-392-67364	Sequence 35280, A
19	21.4	64.8	12414	US-09-620-392-67364	Sequence 35280, A
20	21.4	64.8	12414	US-09-620-392-67364	Sequence 35280, A
21	21.4	64.8	12414	US-09-620-392-67364	Sequence 35280, A
22	21.4	64.8	12414	US-09-620-392-67364	Sequence 35280, A
23	21.4	64.8	12414	US-09-620-392-67364	Sequence 35280, A
24	21.4	64.8	12414	US-09-620-392-67364	Sequence 35280, A
25	21.4	64.8	12414	US-09-620-392-67364	Sequence 35280, A
26	21.4	64.8	12414	US-09-620-392-67364	Sequence 35280, A
27	21.4	64.8	12414	US-09-620-392-67364	Sequence 35280, A
28	21.4	64.8	12414	US-09-620-392-67364	Sequence 35280, A
29	21.4	64.8	12414	US-09-620-392-67364	Sequence 35280, A
30	21.4	64.8	12414	US-09-620-392-67364	Sequence 35280, A
31	21.4	64.8	12414	US-09-620-392-67364	Sequence 35280, A
32	21.4	64.8	12414	US-09-620-392-67364	Sequence 35280, A
33	21.4	64.8	12414	US-09-620-392-67364	Sequence 35280, A
34	21.4	64.8	12414	US-09-620-392-67364	Sequence 35280, A
35	21.4	64.8	12414	US-09-620-392-67364	Sequence 35280, A
36	21.4	64.8	12414	US-09-620-392-67364	Sequence 35280, A
37	21.4	64.8	12414	US-09-620-392-67364	Sequence 35280, A
38	21.4	64.8	12414	US-09-620-392-67364	Sequence 35280, A
39	21.4	64.8	12414	US-09-620-392-67364	Sequence 35280, A
40	21.4	64.8	12414	US-09-620-392-67364	Sequence 35280, A
41	21.4	64.8	12414	US-09-620-392-67364	Sequence 35280, A
42	21.4	64.8	12414	US-09-620-392-67364	Sequence 35280, A
43	21.4	64.8	12414	US-09-620-392-67364	Sequence 35280, A

```
32 20.2 61.2 1244 62 US-60-234-690-977 Sequence 977, App
33 20.2 61.2 1244 71 US-60-323-843-977 Sequence 977, App
34 20.2 61.2 2178 18 US-09-493-276-32 Sequence 32, Appl
35 20.2 61.2 2268 26 US-09-663-779-976 Sequence 576, Appl
36 20.2 61.2 106901 63 US-60-243-733-1 Sequence 1, Appl1
37 20 60.6 225 19 US-09-521-640-237893 Sequence 237893,
38 20 60.6 419 18 US-09-465-877-1396 Sequence 1396, Ap
39 20 60.6 419 34 US-09-906-555-1396 Sequence 1396, Ap
40 20 60.6 427 17 US-09-399-720-9839 Sequence 9839, Ap
41 20 60.6 427 34 US-09-921-378-9839 Sequence 9839, Ap
42 20 60.6 434 18 US-09-465-877-2338 Sequence 2338, Ap
43 20 60.6 434 34 US-09-906-555-2338 Sequence 2338, Ap
44 20 60.6 444 18 US-09-465-877-3009 Sequence 3009, Ap
45 20 60.6 444 19 US-09-519-705-5860 Sequence 5860, Ap
```

ALIGNMENTS

```
RESULT 1
US-09-554-024-1
```

```
; Sequence 1, Application US/09554024
```

```
; GENERAL INFORMATION:
```

```
; APPLICANT: Freysinet, Georges
```

```
; APPLICANT: Derose, Richard
```

```
; APPLICANT: Hoffman, Jules
```

```
; TITLE OF INVENTION: Gene Coding for Thanatin, Vector
```

```
; TITLE OF INVENTION: Containing Same and Resulting Transformed Disease-Resistant
```

```
; TITLE OF INVENTION: Plants
```

```
; FILE REFERENCE: A33207-PCT-USA
```

```
; CURRENT APPLICATION NUMBER: US/09/554,024
```

```
; PRIOR FILING DATE: 2000-05-08
```

```
; PRIOR APPLICATION NUMBER: PCT/FR98/02375
```

```
; PRIOR FILING DATE: 1998-11-06
```

```
; PRIOR APPLICATION NUMBER: FR 97/14,263
```

```
; PRIOR FILING DATE: 1997-11-07
```

```
; NUMBER OF SEQ ID NOS: 14
```

```
; SOFTWARE: FastSeq for Windows Version 3.0
```

```
; SEQ ID NO 1
```

```
; LENGTH: 33
```

```
; TYPE: DNA
```

```
; ORGANISM: Psodius maculiventis
```

```
; FEATURE:
```

```
; NAME/KEY: CDS
```

```
; LOCATION: (1)...(33)
```

```
US-09-554-024-1
```

```
Query Match 100.0%; Score 33; DB 22; Length 33;
```

```
Best Local Similarity 100.0%; Pred. No. 0.00075;
```

```
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1 atcattactgcaacagagagactgtaagtgc 33
```

```
Db 1 atcattactgcaacagagagactgtaagtgc 33
```

```
RESULT 2
US-09-554-024-2
```

```
; Sequence 2, Application US/09554024
```

```
; GENERAL INFORMATION:
```

```
; APPLICANT: Freysinet, Georges
```

```
; APPLICANT: Derose, Richard
```

```
; APPLICANT: Hoffman, Jules
```

```
; TITLE OF INVENTION: Gene Coding for Thanatin, Vector
```

```
; TITLE OF INVENTION: Containing Same and Resulting Transformed Disease-Resistant
```

```
; TITLE OF INVENTION: Plants
```

```
; FILE REFERENCE: A33207-PCT-USA
```

```
; CURRENT APPLICATION NUMBER: US/09/554,024
```

```
; PRIOR FILING DATE: 2000-05-08
```

```
; PRIOR APPLICATION NUMBER: PCT/FR98/02375
```

```
; PRIOR FILING DATE: 1998-11-06
```

```
; PRIOR APPLICATION NUMBER: FR 97/14,263
```

```
; PRIOR FILING DATE: 1997-11-07
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 63
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(63)
; OTHER INFORMATION: Derived from Psodius maculiventis
US-09-554-024-2
```

```
Query Match 100.0%; Score 33; DB 22; Length 63;
```

```
Best Local Similarity 100.0%; Pred. No. 0.00085;
```

```
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1 atcattactgcaacagagagactgtaagtgc 33
Db 22 atcattactgcaacagagagactgtaagtgc 54
```

```
RESULT 3
```

```
US-09-554-024-9/c
```

```
; Sequence 9, Application US/09554024
```

```
; GENERAL INFORMATION:
```

```
; APPLICANT: Freysinet, Georges
```

```
; APPLICANT: Derose, Richard
```

```
; APPLICANT: Hoffman, Jules
```

```
; TITLE OF INVENTION: Gene Coding for Thanatin, Vector
```

```
; TITLE OF INVENTION: Containing Same and Resulting Transformed Disease-Resistant
```

```
; TITLE OF INVENTION: Plants
```

```
; FILE REFERENCE: A33207-PCT-USA
```

```
; CURRENT APPLICATION NUMBER: US/09/554,024
```

```
; PRIOR FILING DATE: 2000-05-08
```

```
; PRIOR APPLICATION NUMBER: PCT/FR98/02375
```

```
; PRIOR FILING DATE: 1998-11-06
```

```
; PRIOR APPLICATION NUMBER: FR 97/14,263
```

```
; PRIOR FILING DATE: 1997-11-07
```

```
; NUMBER OF SEQ ID NOS: 14
```

```
; SOFTWARE: FastSeq for Windows Version 3.0
```

```
; SEQ ID NO 9
```

```
; LENGTH: 87
```

```
; TYPE: DNA
```

```
; ORGANISM: Artificial Sequence
```

```
; FEATURE:
```

```
; OTHER INFORMATION: Synthetic oligonucleotide
```

```
US-09-554-024-9
```

```
Query Match 100.0%; Score 33; DB 22; Length 87;
```

```
Best Local Similarity 100.0%; Pred. No. 0.0009;
```

```
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1 atcattactgcaacagagagactgtaagtgc 33
```

```
Db 77 ATCATCTACTGCACAGAGAGACTGCTAAGTGC 45
```

```
RESULT 4
US-09-554-024-3
```

```
; Sequence 3, Application US/09554024
```

```
; GENERAL INFORMATION:
```

```
; APPLICANT: Freysinet, Georges
```

```
; APPLICANT: Derose, Richard
```

```
; APPLICANT: Hoffman, Jules
```

```
; TITLE OF INVENTION: Gene Coding for Thanatin, Vector
```

```
; TITLE OF INVENTION: Containing Same and Resulting Transformed Disease-Resistant
```

```
; TITLE OF INVENTION: Plants
```

```
; FILE REFERENCE: A33207-PCT-USA
```

```
; CURRENT APPLICATION NUMBER: US/09/554,024
```

```
; PRIOR FILING DATE: 2000-05-08
```



```
; PRIOR APPLICATION NUMBER: PCT/FR98/02375
; PRIOR FILING DATE: 1998-11-06
; PRIOR APPLICATION NUMBER: FR 97/14,263
; PRIOR FILING DATE: 1997-11-07
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 98
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Derived from Psodius maculiventris
; NAME/KEY: CDS
; LOCATION: (1)...(63)
US-09-554-024-3

Query Match          100.0%; Score 33; DB 22; Length 98;
Best Local Similarity 100.0%; Pred. No. 0.00093;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atcatctactgcacacagagagactggtagtgc 33
    |||
Db . 22 atcatctactgcacacagagagactggtagtgc 54

RESULT 5
; Sequence 5, Application US/09554024
; GENERAL INFORMATION:
; APPLICANT: Freyssinet, Georges
; APPLICANT: Derose, Richard
; APPLICANT: Hoffman, Jules
; TITLE OF INVENTION: Gene Coding for Thnatin, Vector
; TITLE OF INVENTION: Containing Same and Resulting Transformed Disease-Resistant
; FILE REFERENCE: A33207-PCT-USA
; CURRENT APPLICATION NUMBER: US/09/554,024
; CURRENT FILING DATE: 2000-05-08
; PRIOR APPLICATION NUMBER: PCT/FR98/02375
; PRIOR FILING DATE: 1998-11-06
; PRIOR APPLICATION NUMBER: FR 97/14,263
; PRIOR FILING DATE: 1997-11-07
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 197
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Derived from Psodius maculiventris
; NAME/KEY: CDS
; LOCATION: (12)...(164)
US-09-554-024-5

Query Match          100.0%; Score 33; DB 22; Length 197;
Best Local Similarity 100.0%; Pred. No. 0.0011;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atcatctactgcacacagagagactggtagtgc 33
    |||
Db 123 atcatctactgcacacagagagactggtagtgc 155

RESULT 6
; Sequence 5, Application US/08265534A
; GENERAL INFORMATION:
; APPLICANT: Yaver, Debbie Sue
; APPLICANT: Xu, Feng
; APPLICANT: Dalpoge, Henrik
; APPLICANT: Schneider, Palie
```

```
; TITLE OF INVENTION: PURIFIED POLYPORUS LACCASES AND
; TITLE OF INVENTION: NUCLEIC ACIDS ENCODING SAME
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Novo Nordisk of North America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/265,534A
; FILING DATE: 24-JUN-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lowmey, Karen A.
; REGISTRATION NUMBER: 31,274
; REFERENCE/DOCKET NUMBER: 4185,000-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3102 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Polyporus pinistius
; FEATURE:
; NAME/KEY: intron
; LOCATION: 666..720
; FEATURE:
; NAME/KEY: intron
; LOCATION: 790..845
; FEATURE:
; NAME/KEY: intron
; LOCATION: 1125..1182
; FEATURE:
; NAME/KEY: intron
; LOCATION: 1390..1450
; FEATURE:
; NAME/KEY: intron
; LOCATION: 1607..1661
; FEATURE:
; NAME/KEY: intron
; LOCATION: 1863..1918
; FEATURE:
; NAME/KEY: intron
; LOCATION: 1976..2025
; FEATURE:
; NAME/KEY: intron
; LOCATION: 2227..2285
; FEATURE:
; NAME/KEY: intron
; LOCATION: 2403..2458
; FEATURE:
; NAME/KEY: intron
; LOCATION: 2576..2627
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join (665..721, 789..846, 1124..1183, 1389..1451,
; LOCATION: 1606..1662, 1862..1919, 1975..2026, 2226..2286, 2402..2459,
; LOCATION: 2575..2628).
```

US-08-265-534A-5

Query Match 67.9%: Score 22.4; DB 6; Length 3102;
Best Local Similarity 81.2%; Pred. No. 72;
Matches 26; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 2 tcatctactgcacagagagactgtaagtgc 33
Db 2623 TCATCTTTGCACAGAGAGACTGGAACAAGC 2654

RESULT 7
US-08-456-341-5
; Sequence 5, Application US/08456341
; GENERAL INFORMATION:
; APPLICANT: Yaver, Debbie Sue
; APPLICANT: Xu, Feng
; APPLICANT: Dalboge, Henrik
; APPLICANT: Schneider, Palie
; TITLE OF INVENTION: PURIFIED POLYPORUS LACCASES AND
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Novo Nordisk of North America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/456,341
; FILING DATE: 01-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/265,534
; FILING DATE: 24-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Lowney, Karen A.
; REGISTRATION NUMBER: 31,274
; REFERENCE/DOCKET NUMBER: 4185,000-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0298
; TELEFAX: 212 867 0123
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3102 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Polyporus plustus
; FEATURE:
; NAME/KEY: Intron
; LOCATION: 666..720
; FEATURE:
; NAME/KEY: Intron
; LOCATION: 790..845
; FEATURE:
; NAME/KEY: Intron
; LOCATION: 1125..1182
; FEATURE:
; NAME/KEY: Intron
; LOCATION: 1390..1450
; FEATURE:
; NAME/KEY: Intron
; LOCATION: 1607..1661
; FEATURE:
; NAME/KEY: Intron
; LOCATION: 1863..1918

FEATURE:
; NAME/KEY: Intron
; LOCATION: 1976..2025
; FEATURE:
; NAME/KEY: Intron
; LOCATION: 2227..2285
; FEATURE:
; NAME/KEY: Intron
; LOCATION: 2403..2458
; FEATURE:
; NAME/KEY: Intron
; LOCATION: 2576..2627
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1606..1662, 1862..1919, 1975..2026, 2226..2286, 2402..2459,
; LOCATION: 2575..2628).

Query Match 67.9%: Score 22.4; DB 8; Length 3102;
Best Local Similarity 81.2%; Pred. No. 72;
Matches 26; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 2 tcatctactgcacagagagactgtaagtgc 33
Db 2623 TCATCTTTGCACAGAGAGACTGGAACAAGC 2654

RESULT 8
US-08-457-001-5
; Sequence 5, Application US/08457001
; GENERAL INFORMATION:
; APPLICANT: Yaver, Debbie Sue
; APPLICANT: Xu, Feng
; APPLICANT: Dalboge, Henrik
; APPLICANT: Schneider, Palie
; TITLE OF INVENTION: PURIFIED POLYPORUS LACCASES AND
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Novo Nordisk of North America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/457,001
; FILING DATE: 01-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/265,534
; FILING DATE: 24-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Lowney, Karen A.
; REGISTRATION NUMBER: 31,274
; REFERENCE/DOCKET NUMBER: 4185,000-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0298
; TELEFAX: 212 867 0123
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3102 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)


```

1  APPLICANT: Yaver, Debbie Sue
2  APPLICANT: Xu, Feng
3  APPLICANT: Dalboge, Henrik
4  APPLICANT: Schneider, Palle
5  TITLE OF INVENTION: PURIFIED POLYPORUS LACCASES AND
6  TITLE OF INVENTION: NUCLEIC ACIDS ENCODING SAME
7  NUMBER OF SEQUENCES: 10
8  CORRESPONDENCE ADDRESS:
9  ADDRESSEE: Novo Nordisk of North America, Inc.
10 STREET: 405 Lexington Avenue, Suite 6400
11 CITY: New York
12 STATE: New York
13 COUNTRY: U.S.A.
14 ZIP: 10174-6401
15 COMPUTER READABLE FORM:
16 MEDIUM TYPE: Floppy disk
17 COMPUTER: IBM PC compatible
18 OPERATING SYSTEM: PC-DOS/MS-DOS
19 SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
20 CURRENT APPLICATION DATA:
21 APPLICATION NUMBER: US/08/457,543
22 FILING DATE: 01-JUN-1995
23 CLASSIFICATION: 435
24 PRIOR APPLICATION DATA:
25 APPLICATION NUMBER: 08/265,534
26 FILING DATE: 24-JUN-1994
27 ATTORNEY/AGENT INFORMATION:
28 NAME: Lowney, Karen A.
29 REGISTRATION NUMBER: 31,274
30 REFERENCE/DOCKET NUMBER: 4185.000-US
31 TELECOMMUNICATION INFORMATION:
32 TELEPHONE: 212 867 0123
33 TELEFAX: 212 867 0298
34 INFORMATION FOR SEQ ID NO: 5:
35 SEQUENCE CHARACTERISTICS:
36 LENGTH: 3102 base pairs
37 TYPE: nucleic acid
38 STRANDEDNESS: double
39 TOPOLOGY: linear
40 MOLECULE TYPE: DNA (genomic)
41 ORIGINAL SOURCE:
42 ORGANISM: Polyporus pluitus
43 FEATURE:
44 NAME/KEY: Intron
45 LOCATION: 666..720
46 FEATURE:
47 NAME/KEY: Intron
48 LOCATION: 790..845
49 FEATURE:
50 NAME/KEY: Intron
51 -LOCATION: 1125..1182
52 FEATURE:
53 NAME/KEY: Intron
54 LOCATION: 1390..1450
55 FEATURE:
56 NAME/KEY: Intron
57 LOCATION: 1607..1661
58 FEATURE:
59 NAME/KEY: Intron
60 LOCATION: 1863..1918
61 FEATURE:
62 NAME/KEY: Intron
63 LOCATION: 1976..2025
64 FEATURE:
65 NAME/KEY: Intron
66 LOCATION: 2227..2285
67 FEATURE:
68 NAME/KEY: Intron
69 LOCATION: 2403..2458
70 FEATURE:
71 NAME/KEY: Intron
72 LOCATION: 2576..2627
73 FEATURE:

```

```

;      NAME/KEY:   CDS
;      LOCATION:   join (665..721, 789..846, 1124..1183, 1389..1451,
;      LOCATION:   1606..1662, 1862..1919, 1975..2026, 2226..2286, 2402..2459
;      LOCATION:   2575..2628) .
;
US-08-457-543-5

```

Query Match	67.9%;	Score 22.4;	DB 8;	Length 3102;
Best Local Similarity	81.2%;	Pred. No. 72;		
Matches 26; Conservative	0;	Mismatches 6;	Indels 0;	Gaps 0;

QY	2	tcatctactgcacacagcagcagctgtaagtgc	33
Db	2623	TCATCTTTTGCAACAGGAGGACTGGAACAAGC	2654

RESULT 11
US-09-248-796-8333/C
; Sequence 8333, Application US/09248796

? APPLICANT: Keith Weinstein et al
 ? TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBIDUS
 ? TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
 ? FILE REFERENCE: 107196.132
 ? CURRENT APPLICATION NUMBER: US/09/248.796
 ? CURRENT FILING DATE: 1999-02-12
 ? NUMBER OF SEQ ID NOS: 28206
 ? SEQ ID NO 8333
 ? LENGTH: 1230
 ? TYPE: DNA
 ? ORGANISM: *Candida albicans*
 ? US-09-248-796-8333

Query Match	66.7%;	Score 22;	DB 16;	Length 1230;
Best Local Similarity	83.3%;	Pred. No. 89;		
Matches 25; Conservative	0;	Mismatches 5;	Indels 0;	Gaps 0;

Qy	3	catctactgcaacagagagactggttaagt	32
Db	625	CATGTACTGCACACAGGATGACTCGTCATG	596

```

RESULT      12
US-60-096-409-8333/C
; Sequence 8333, Application US/60096409A
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBI
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-016P
; CURRENT APPLICATION NUMBER: US/60/096,409A
; CURRENT FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28206
; SEQ ID NO 8353
; LENGTH: 1230
; TYPE: DNA
; ORGANISM: Candida albicans
US-60-096-409-8333

```

Query Match	66.7%;	Score 22;	DB 48;	Length 1230;
Best Local Similarity	83.3%;	Pred. No. 89;		
Matches 25; Conservative	0;	Mismatches 5;	Indels 0;	Gaps 0;

OY		3	catctactgcacacaggaagtgtgtaagt	32
Db	625	CATGTACTGCACAAGGATGACCTCGTCATG	596	

RESULT 13
US-05-620-392-35280/c
: Sequence 35280, Application US/09620392

```

: GENERAL INFORMATION:
: APPLICANT: Boukharov, Andrey A.
: APPLICANT: Kovalic, David K.
: APPLICANT: Liu, Jingdong
: APPLICANT: McIninch, James
: TITLE OF INVENTION: Plant Genome Sequence And Uses Thereof
: FILE REFERENCE: 38-21(51237)E
: CURRENT APPLICATION NUMBER: US/09/620.392
: CURRENT FILING DATE: 2000-07-19
: NUMBER OF SEQ ID NOS: 69652
: SEQ ID NO 35280
: LENGTH: 12414
: TYPE: DNA
: ORGANISM: Oryza sativa
: US-09-620-392-35280

```

	Query Match	64.8%	Score 21.4	DB 24:	Length 12414;
	Best Local Similarity	80.6%	Pred. No. 2.6e+02;		
	Matches 25; Conservative	0;	Mismatches 6;	Indels 0;	Gaps 0;
Oy	1 atcattctactgcacacaggaagacgttgtaagt	31			
Db	3272 ATAAATATTCTCAACAGGAGGTGGTTACT	3242			

```

RESULT 14
US-09-620-392-67364/c
: Sequence 67364, Application US/09620392
: GENERAL INFORMATION:
: APPLICANT: Soukharov, Andrey A.
: APPLICANT: Kovalic, David K.
: APPLICANT: Liu, Jingdong
: APPLICANT: Mcclinch, James
: TITLE OF INVENTION: Plant Genome Sequence And Uses Thereof
: FILE REFERENCE: 38-21(51237)E
: CURRENT APPLICATION NUMBER: US/09/620.392
: CURRENT FILING DATE: 2000-07-19
: NUMBER OF SEQ ID NOS: 69652
: SEQ ID NO 67364
: LENGTH: 21712
: TYPE: DNA
: ORGANISM: Oryza sativa
: US-09-620-392-67364

```

	Query Match	64.8%	Score 21.4	DB 22	Length 21112
	Best Local Similarity	80.6%	Pred. No. 2.9e02		
	Matches 25; Conservative	0;	Mismatches 6;	Indels 0;	Gaps 0;
OY	1 atcattctactgcacagaggactcgtaagt	31			
Db	3359 ATAAATATACTGCAACAGGAGGTGGTTAAGT	3359			

```

RESULT 15
US-09-702-134-7465/c
: Sequence 7465, Application US/09702134
: GENERAL INFORMATION:
: APPLICANT: Boukharov, Andrey A.
: APPLICANT: Cao, Yongwei
: APPLICANT: Kovalic, David K.
: APPLICANT: Liu, Jingdong
: APPLICANT: McIninch, James
: APPLICANT: Wu, Wei
: TITLE OF INVENTION: Plant Genome Sequence and Uses Thereof
: FILE REFERENCE: 38-21(51237)F
: CURRENT APPLICATION NUMBER: US/09/702,134
: CURRENT FILING DATE: 2000-10-31
: NUMBER OF SEQ ID NOS: 52202
: SEQ ID NO 7465
: LENGTH: 66687
: TYPE: DNA

```

ORGANISM: *Oryza sativa*
US-09-702-134-7465

Query Match	64.8%	Score 21.4	DB 28	Length 66687
Best Local Similarity	80.6%	Pred. No. 3,66+02		
Matches 25; Conservative	0;	Mismatches 6;	Indels 0;	Gaps 0;
Qy	1	atcattctctcgcacacgagagactcgttaagt	31	
Db	3349	ATAATATATCTCTCAACAGCGGTGGTTAAGT	3319	

```
Search completed: May 12, 2002, 21:33:35
Job time: 9231 sec
```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 12, 2002, 18:14:43 ; Search time 76.1 Seconds
(without alignments)
106.516 Million cell updates/sec

Title: US-09-554-024-1
Perfect score: 33
Sequence: 1 atcattctgcaacagagagactgttaagtc 33

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PTOS.COMB.seq:*
6: /cgn2_6/ptodata/2/ina/Dackfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	length DB	ID	Description
1	22.4	67.9	3102	1	US-08-462-484-5
2	22.4	67.9	3102	1	US-08-441-147-5
3	22.4	67.9	3102	5	PCT-US95-07536-5
4	21	63.6	234	1	US-08-358-160-143
5	20.2	61.2	18073	4	US-09-078-294-12
6	19.8	60.0	180	1	US-08-358-160-145
7	18.8	57.0	824	1	US-08-158-353-1
8	18.6	56.4	330	5	PCT-US91-02766-13
9	18.6	56.4	375	5	PCT-US91-02766-15
10	18.6	56.4	385	6	5428135-3
11	18.6	56.4	649	6	5194396-12
12	18.6	56.4	649	6	5219739-12
13	18.6	56.4	748	1	US-08-387-845-1
14	18.6	56.4	748	2	US-08-778-275-1
15	18.6	56.4	748	3	US-08-867-352-1
16	18.6	56.4	1247	5	PCT-US91-02766-19
17	18.6	56.4	1247	6	5219759-3
18	18.6	56.4	1316	5	PCT-US91-02766-21
19	18	54.5	18609	4	US-08-943-731-1
20	17.6	53.3	924	3	US-08-983-409-3
21	17.4	52.7	600	4	US-09-364-083-1
22	17.4	52.7	639	4	US-09-196-293-10
23	17.4	52.7	639	4	US-08-209-603E-10
24	17.4	52.7	639	4	US-08-233-836C-33
25	17.4	52.7	949	5	PCT-US91-08254-1
26	17.4	52.7	949	5	PCT-US91-08254-2
27	17.4	52.7	1611	1	US-08-061-062A-5

28	17.4	52.7	1611	3	US-08-536-150-5	Sequence 5, Appl
29	17.4	52.7	3250	1	US-08-061-062A-7	Sequence 7, Appl
30	17.4	52.7	3250	3	US-08-536-150-7	Sequence 7, Appl
31	17.4	52.7	5682	2	US-08-663-998-4	Sequence 4, Appl
32	17.2	52.1	31	1	US-08-195-874-3	Sequence 3, Appl
33	17.2	52.1	31	1	US-08-195-874-8	Sequence 8, Appl
34	17.2	52.1	31	5	PCT-US95-01671-3	Sequence 3, Appl
35	17.2	52.1	31	5	PCT-US95-01671-8	Sequence 8, Appl
36	17.2	52.1	2280	2	US-09-055-097-4	Sequence 4, Appl
37	17.2	52.1	7898	4	US-08-984-709A-49	Sequence 49, Appl
38	17	51.5	233	1	US-08-444-142-3	Sequence 3, Appl
39	17	51.5	233	1	US-08-444-131-3	Sequence 3, Appl
40	17	51.5	777	4	US-08-998-416-218	Sequence 218, App
41	17	51.5	1180	2	US-08-867-087B-16	Sequence 16, Appl
42	17	51.5	1316	6	5219759-1	Patent No. 5219759
43	17	51.5	1987	1	US-07-999-280A-23	Sequence 23, Appl
44	17	51.5	1987	1	US-08-426-036-23	Sequence 23, Appl
45	17	51.5	1987	1	US-08-426-279-23	Sequence 23, Appl

ALIGNMENTS

RESULT 1
US-08-462-484-5
Sequence 5, Application US/08462484
Patent No. 5667531
GENERAL INFORMATION:
APPLICANT: Yaver, Debbie Sue
APPLICANT: Xu, Feng
APPLICANT: Dalboge, Henrik
APPLICANT: Schneider, Palie
APPLICANT: Aaslyng, Dorrit A.
TITLE OF INVENTION: PURIFIED POLYPORUS LACCASES AND
TITLE OF INVENTION: NUCLEIC ACIDS ENCODING SAME
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 56675310 No. 5667531disk of No. 5667531th America, Inc.
STREET: 405 Lexington Avenue, Suite 6400
City: New York
State: New York
COUNTRY: U.S.A.
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25 (ERO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/462,484
FILING DATE: 05-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/441,147
FILING DATE: 15-MAY-1995
ATTORNEY/AGENT INFORMATION:
NAME: Lowney, Karen A.
REGISTRATION NUMBER: 31,274
REFERENCE/DOCKET NUMBER: 4185, 010-US-
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 867 0123
TELEFAX: 212 878 9655
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 3102 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Polyporus plinthus
FEATURE:
NAME/KEY: Intron

```
LOCATION: 666..720
FEATURE:
NAME/KEY: Intron
LOCATION: 790..845
FEATURE:
NAME/KEY: Intron
LOCATION: 1125..1182
FEATURE:
NAME/KEY: Intron
LOCATION: 1390..1450
FEATURE:
NAME/KEY: Intron
LOCATION: 1607..1661
FEATURE:
NAME/KEY: Intron
LOCATION: 1863..1918
FEATURE:
NAME/KEY: Intron
LOCATION: 1976..2025
FEATURE:
NAME/KEY: Intron
LOCATION: 2227..2285
FEATURE:
NAME/KEY: Intron
LOCATION: 2403..2458
FEATURE:
NAME/KEY: Intron
LOCATION: 2576..2627
FEATURE:
NAME/KEY: CDS
LOCATION: 1016 (665..721, 789..846, 1124..1183, 1389..1451,
1606..1662, 1862..1919, 1975..2026, 2226..2286, 2402..2459,
2575..2628).
US-08-462-484-5
```

```
Query Match      67.9%; Score 22.4; DB 1; Length 3102;
Best Local Similarity 81.2%; Pred. No. 0.83;
Matches 26; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
```

```
QY 2 tcactctacgcacagaggagactggtagtgc 33
||||| ||||||| ||||| ||
Db 2623 TCATCTTTTGCACAGGAGGACTGGAACAAGC 2654
```

```
RESULT 2
US-08-441-147-5
Sequence 5, Application US/08441147
Patent No. 5770418
GENERAL INFORMATION:
APPLICANT: Yaver, Debbie Sue
APPLICANT: Xu, Feng
APPLICANT: Dalboge, Henrik
APPLICANT: Schneider, Palte
APPLICANT: Aslyng, Dorrit A.
TITLE OF INVENTION: PURIFIED POLYPORUS LACCASES AND
NUCLEIC ACIDS ENCODING SAME
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESS: No. 5770418b No. 5770418disk of No. 5770418th America, Inc.
STREET: 405 Lexington Avenue, Suite 6400
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/441,147
FILING DATE: 15-MAY-1995
```

```
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Lowney, Karen A.
REGISTRATION NUMBER: 31,274
REFERENCE/DOCKET NUMBER: 4185, 010-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 867 0123
TELEFAX: 212 878 9655
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 3102 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Polyporus plinthus
FEATURE:
NAME/KEY: Intron
LOCATION: 666..720
FEATURE:
NAME/KEY: Intron
LOCATION: 790..845
FEATURE:
NAME/KEY: Intron
LOCATION: 1125..1182
FEATURE:
NAME/KEY: Intron
LOCATION: 1390..1450
FEATURE:
NAME/KEY: Intron
LOCATION: 1607..1661
FEATURE:
NAME/KEY: Intron
LOCATION: 1863..1918
FEATURE:
NAME/KEY: Intron
LOCATION: 1976..2025
FEATURE:
NAME/KEY: Intron
LOCATION: 2227..2285
FEATURE:
NAME/KEY: Intron
LOCATION: 2403..2458
FEATURE:
NAME/KEY: Intron
LOCATION: 2576..2627
FEATURE:
NAME/KEY: CDS
LOCATION: 1016 (665..721, 789..846, 1124..1183, 1389..1451,
1606..1662, 1862..1919, 1975..2026, 2226..2286, 2402..2459,
2575..2628).
US-08-441-147-5
```

```
Query Match      67.9%; Score 22.4; DB 1; Length 3102;
Best Local Similarity 81.2%; Pred. No. 0.83;
Matches 26; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
```

```
QY 2 tcactctacgcacagaggagactggtagtgc 33
||||| ||||||| ||||| ||
Db 2623 TCATCTTTTGCACAGGAGGACTGGAACAAGC 2654
```

```
RESULT 3
PCT-US95-07536-5
Sequence 5, Application PC/TUS9507536
GENERAL INFORMATION:
APPLICANT:
APPLICANT:
TITLE OF INVENTION: PURIFIED POLYPORUS LACCASES AND
NUCLEIC ACIDS ENCODING SAME
NUMBER OF SEQUENCES: 10
```


CORRESPONDENCE ADDRESS:
 ADDRESSEE: Novo Nordisk of North America, Inc.
 STREET: 405 Lexington Avenue, Suite 6400
 COUNTRY: U.S.A.
 ZIP: 10174-6401
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US95/07536
 FILING DATE: 15-June-1995
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/265,534
 FILING DATE: 24-June-1994
 ATTORNEY/AGENT INFORMATION:
 NAME: Lowmey, Karen A.
 REGISTRATION NUMBER: 31,274
 REFERENCE/DOCKET NUMBER: 4185.204-WO
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 212 867 0123
 TELEFAX: 212 878 9655
 INFORMATION FOR SEQ ID NO: 5:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 3102 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 ORIGINAL SOURCE:
 ORGANISM: Polyporus pluvius
 FEATURE:
 NAME/KEY: Inttron
 LOCATION: 666..720
 FEATURE:
 NAME/KEY: Inttron
 LOCATION: 790..845
 FEATURE:
 NAME/KEY: Inttron
 LOCATION: 1125..1182
 FEATURE:
 NAME/KEY: Inttron
 LOCATION: 1390..1450
 FEATURE:
 NAME/KEY: Inttron
 LOCATION: 1607..1661
 FEATURE:
 NAME/KEY: Inttron
 LOCATION: 1863..1918
 FEATURE:
 NAME/KEY: Inttron
 LOCATION: 1976..2025
 FEATURE:
 NAME/KEY: Inttron
 LOCATION: 2227..2285
 FEATURE:
 NAME/KEY: Inttron
 LOCATION: 2403..2458
 FEATURE:
 NAME/KEY: Inttron
 LOCATION: 2576..2627
 NAME/KEY:
 LOCATION: join (665..721, 789..846, 1124..1183, 1389..1451,
 1606..1662, 1862..1919, 1975..2026, 2226..2286, 2402..2459,
 2575..2628).
 PCT-US95-07536-5

Query Match 67.9%; Score 22.4; DB 5; Length 3102;
 Best Local Similarity 81.2%; Pred. No. 0.83;
 Matches 20; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Oy 2 tcactactgcacagagactgtaagtc 33
 Db 2623 TCATCTTTGCACAGAGACTGGACACAGC 2654
 RESULT 4
 US-08-358-160-143
 Sequence 143, Application US/08358160
 Patent No. 5663143
 GENERAL INFORMATION:
 APPLICANT: LEY, Arthur C.
 APPLICANT: LADNER, Robert C.
 APPLICANT: GUTERMAN, Sonia K.
 APPLICANT: ROBERTS, Bruce L.
 APPLICANT: MARKLAND, William
 APPLICANT: KENT, Rachel B.
 TITLE OF INVENTION: ENGINEERED HUMAN-DERIVED KUNITZ
 NUMBER OF SEQUENCES: 234
 NUMBER OF INVENTIONS: DOMAINS THAT INHIBIT HUMAN NEUTROPHIL ELASTASE
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: BROWDY AND NEIMARK
 STREET: 419 Seventh Street, N.W. Suite 300
 CITY: Washington
 STATE: District of Columbia
 COUNTRY: USA
 ZIP: 20004
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/358,160
 FILING DATE: 16-DEC-1994
 CLASSIFICATION: 514
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/133,031
 FILING DATE: 13-OCT-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/009,319
 FILING DATE: 26-JAN-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/664,989
 FILING DATE: 01-MAR-1991
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/487,063
 FILING DATE: 02-MAR-1990
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/240,160
 FILING DATE: 02-SEP-1988
 ATTORNEY/AGENT INFORMATION:
 NAME: Cooper, Iver P.
 REGISTRATION NUMBER: 28,005
 REFERENCE/DOCKET NUMBER: LEY-1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 202-628-5197
 TELEFAX: 202-737-3528
 TELEX: 248633
 INFORMATION FOR SEQ ID NO: 143:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 234 bases
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: other nucleic acid
 DESCRIPTION: Synthetic DNA fragment
 US-08-358-160-143

Query Match 63.6%; Score 21; DB 1; Length 234;
 Best Local Similarity 82.8%; Pred. No. 1.9;
 Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 5 tctactcacaagagagactgtaagtc 33
||||| ||||| ||||| ||||| |||||
DB 116 TCTACTACACTCGAGATTGTGTAAGTC 144

RESULT 5
US-09-078-294-12/c
Sequence 12, Application US/09078294
Patent No. 6265211
GENERAL INFORMATION:
APPLICANT: Choo, Kong-Hong Andy
APPLICANT: Du Sart, Desiree
APPLICANT: Cancilla, Michael R.
TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
FILE REFERENCE: Davies Col
CURRENT APPLICATION NUMBER: US/09/078,294
CURRENT FILING DATE: 1998-05-13
NUMBER OF SEQ ID NOS: 29
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 12
LENGTH: 18073
TYPE: DNA
ORGANISM: BAC-F2 contlg 8
US-09-078-294-12

Query Match 61.2%; Score 20.2; DB 4; Length 18073;
Best Local Similarity 88.0%; Pred. No. 11;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 8 actgcacacagagagactgtaagtc 32
||||| ||||| ||||| ||||| |||||
DB 8535 ACTGTTACAGAGAGACTGTAAATG 8511

RESULT 6
US-08-358-160-145
Sequence 145, Application US/08358160
Patent No. 5663143
GENERAL INFORMATION:
APPLICANT: LEY, Arthur C.
APPLICANT: LADNER, Robert C.
APPLICANT: GUTERMAN, Sonia K.
APPLICANT: ROBERTS, Bruce L.
APPLICANT: MARKLAND, William
APPLICANT: KENT, Rachel B.
TITLE OF INVENTION: ENGINEERED HUMAN-DERIVED KUNITZ
TITLE OF INVENTION: DOMAINS THAT INHIBIT HUMAN NEUTROPHIL ELASTASE
NUMBER OF SEQUENCES: 234
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEWMARK
STREET: 419 Seventh Street, N.W. Suite 300
CITY: Washington
STATE: District of Columbia
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/358,160
FILING DATE: 16-DEC-1994
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/133,031
FILING DATE: 13-OCT-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/009,319
FILING DATE: 26-JAN-1993
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/664,989
FILING DATE: 01-MAR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/487,063
FILING DATE: 02-MAR-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/240,160
FILING DATE: 02-SEP-1988
ATTORNEY/AGENT INFORMATION:
NAME: Cooper, Iver P.
REGISTRATION NUMBER: 28,005
REFERENCE/DOCKET NUMBER: LEY=1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
TELEX: 248633
INFORMATION FOR SEQ ID NO: 145:
SEQUENCE CHARACTERISTICS:
LENGTH: 180 bases
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: synthetic DNA fragment
US-08-358-160-145

Query Match 60.0%; Score 19.8; DB 1; Length 180;
Best Local Similarity 75.9%; Pred. No. 5.8;
Matches 22; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 5 tctactcacaagagagactgtaagtc 33
:||||| ||||| ||||| ||||| |||||
DB 62 DSTACTACACTCGAGATTGTGTAAGTC 90

RESULT 7
US-08-158-353-1/c
Sequence 1, Application US/08158353
Patent No. 5620862
GENERAL INFORMATION:
APPLICANT: Padula, Steven J.
TITLE OF INVENTION: Methods for Diagnosing Early Lyme
TITLE OF INVENTION: Disease
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Militia Drive
CITY: Lexington
STATE: MA
COUNTRY: USA
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/158,353
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Carroll, Alice O.
REGISTRATION NUMBER: 33,542
REFERENCE/DOCKET NUMBER: OCT93-05
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-861-6240
TELEFAX: 617-861-9540
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 824 base pairs
TYPE: nucleic acid
STRANDEDNESS: single

Mon May 13 10:17:29 2002

Search completed: May 12, 2002, 20:14:35
Job time: 7192 sec

us-09-554-024-1.rni

Page 8